Colour-based gesture recognition for American Sign Language via Hidden Markov Models

Sara Greenberg — University of Waterloo, ON, Canada
Jennifer Blight — University of Waterloo, ON, Canada
Alexander Wong — University of Waterloo, ON, Canada

Abstract

We present a new approach to gesture recognition for use in a sign language learning environment. This method utilizes inexpensive cloth gloves to alleviate the difficulty of hand detection and to allow for feature creation. Salient colours identify the glove base and fingertip markers, which are then used to extract a hand centroid and a convex hull describing the fingertips for each hand. A Hidden Markov Model is created for each sign, as well as an additional threshold model created from all signs. When a candidate sign is performed, the sign of the HMM that produces the greatest likelihood is matched, provided it also exceeds the threshold model likelihood. Isolated recognition testing of the training library indicated 76% accuracy, and continuous recognition testing showed 60% accuracy.

1   Introduction

The Canadian Association of the Deaf approximates that in 2012, there were 350,000 culturally Deaf and 3.15 million hard of hearing Canadians [1]. Age-related hearing difficulty is common, with 47% of adults aged 60-79 experience some form of hearing loss [2]. In Canada, American Sign Language (ASL) is used by most of the Deaf community, but unfortunately the resources for learning ASL are limited and mostly dependent on classroom learning. The popularity of online foreign language learning tools such as Rosetta Stone and Duolingo indicate that computer aided, interactive tools are effective means of learning spoken languages. A similar approach may be beneficial to signed languages.

Such a tool requires a system capable of complex gesture recognition. Existing work that performs this task includes a translation system by Chai *et al.* [3] that uses the Kinect to translate limited signed sentences into English. The system treats hands as point objects, resulting in an underdetermined problem for signs that have the same hand motion but different finger positioning. The research project CopyCat [4] is an ASL teaching game for children. CopyCat performs sign recognition using Hidden Markov Models, but has a complicated setup requiring calibration of wrist-mounted accelerometers.

A novel approach for vision-based sign recognition is proposed using inexpensive materials and providing the user with an uncomplicated setup. Hand and fingertip markers are used to extract sufficient features to correctly identify signs.

2   Methods

We present a new method for recognizing ASL signs performed while wearing inexpensive cloth gloves. This approach makes use of colour thresholding and a sliding temporal window to extract features, and Hidden Markov Models are implemented to classify the time series data and detect signs.

2.1   Video Input and Preprocessing

An RGB camera is used as an input device. A knit glove comprises the glove base and felt pieces mark the fingertips. Colours were selected so as to be highly salient in the YCbCr colour space. An automated calibration process is used to determine colour threshold values. To the resulting binary image we perform dilation and erosion, followed by several median filters of decreasing kernel size.

2.2   Feature Extraction

The fingertip markers allow for the extraction of a convex hull describing the relative position of the user's fingers. Each hand centroid is calculated from the location of the glove colour adjacent to the fingertip markers.

Fig. 1 demonstrates the feature extraction process for a single video frame. The convex hull shape is converted into a feature vector of the central image moments, which have been normalized and re-positioned about the centroid to be shift- and scale-invariant. A feature vector is calculated for each frame to create a time series matrix of n features by t frames. To further reduce noise, the entire time series is smoothed by averaging the last five feature vectors for each frame.
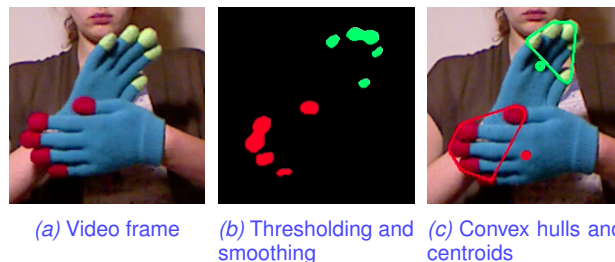

(a) Video frame   *(b)* Thresholding and smoothing   *(c)* Convex hulls and centroids

Fig. 1: Feature extraction steps

2.3   Classification

We classify features by creating a Hidden Markov Model (HMM) [5] for each candidate sign and selecting the model that has the highest probability of generating the given feature set.

To create these models, a training library of data samples was recorded using the Kinect 1.0 with the Fakenect utility. Each sample consists of a subject performing one of ten signs while wearing the gloves. The training library consists of twenty-two samples for each sign, performed by eight individual users.

The start, transition, and emission probabilities are estimated using a Viterbi [6] algorithm. The remaining model parameters are chosen by randomly generating sets of parameters and training models for each. The classifier with the best performance is selected for use.

To avoid always detecting signs, we implement an adaptive threshold above which the classifier likelihood must score in order for the sign to be accepted. This threshold is calculated using a model based on research by Lee and Kim [7] and is a HMM composed of the states of all ten sign models. If the original classifier scores higher than the threshold model, the sign is accepted.

3   Results and Discussion

An isolated recognition test of the training library was performed using k-fold cross validation (*k=5*). The isolated recognition accuracy of the classifier was 76%.

In the continuous recognition test, a user performed each of the candidate signs five times. An observer recorded correct classifications, misclassifications, and false positives. The continuous recognition accuracy of the classifier was 60%, and the most significant source of error was due to the classifier failing to recognize a sign, suggesting that a more relaxed adaptive threshold may increase accuracy.

This method has potential as an affordable sign recognition system. Future works include the addition of adaptive colour thresholding to accommodate variations in lighting conditions, and researching the feasibility of extracting the centroid and convex hull features without the use of gloves.

References

[1] Canadian Association of the Deaf, "Statistics on deaf canadians," http://www.cad.ca/, (23 July 2012).

[2] Statistics Canada, "Hearing loss of canadians, 2012 and 2013," http://www.statcan.gc.ca/pub/82-625-x/2015001/article/14156-eng.htm, (15 April 2015).

[3] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X. Chen, and M. Zhou, "Sign language recognition and translation with kinect," 2013.

[4] Z. Zafrulla, H. Brashear, P. Yin, P. Presti, T. Starner, and H. Hamilton, "American sign language phrase verification in an educational game for deaf children," in *ICPR*, 2010, pp. 3846–3849.

[5] L. R. Rabiner, Juang, and B. H, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, no. January, pp. 257–286, 1986.

[6] J. Forney, G.D., "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 302–309, 1973.

[7] H.-K. Lee and J.-H. Kim, "An hmm-based threshold model approach for gesture recognition," *PAMI*, vol. 21, no. 10, pp. 961–973, 1999.