

Indian Sign Language Interpreter with Android Implementation

Shanmukha Swamy

M N, Ph.D

Professor, Dept. of Electronics
and Communication
SJCE, Mysore
Karnataka, India

Chethan M P

Dept. of Electronics and
Communication
SJCE, Mysore
Karnataka, India

Mahantesh Gatwadi

Dept. of Electronics and
Communication
SJCE, Mysore
Karnataka, India

ABSTRACT

Sign language is used as a communication medium among deaf and dumb people to convey the message with each other. A person who can talk and hear properly (normal person) cannot communicate with deaf and dumb person unless he/she is familiar with sign language. Same case is applicable when a deaf and dumb person wants to communicate with a normal person or blind person. In order to bridge the gap in communication among deaf and dumb community and normal community, lot of research work has been carried out to automate the process of sign language interpretation with the help of image processing and pattern recognition techniques. This paper proposes optimized approaches of implementing the famous Viola Jones algorithm with LBP features for hand gesture recognition which will recognize Indian sign language gestures in a real time environment. The performance analysis of the proposed approaches is presented along with the experimental results. An optimized algorithm has been implemented in the form of an android application and tested with real time data.

Keywords

Indian Sign Language (ISL), American Sign Language (ASL), Local Binary Pattern (LBP), AdaBoost, SIFT, SURF, Viola-Jones.

1. INTRODUCTION

Sign language is seen as the core distinctive feature that defines a deaf community. The role of sign language recognition systems in the society is to ensure that deaf people have equality of opportunity and full participation in society. Sign language is represented basically by continuously varying different hand shapes and movements by a signer. Sign language recognition is decoding and understanding the information embedded in the hand shapes and converting them to meaning full words. Sign language is a natural language of the deaf community. By developing sign language recognition system a hearing impaired person can easily interact with a normal person at different levels in the society.

The existing technologies for sign to text processing uses specialized sensors or may require the usage of gloves, optical markers based on IR reflection and skin color is considered for image processing hence affected by illumination. These are expensive and/or cumbersome. The solution proposed does not require the user to have any specialized sensors attached to his hand or wear any special gloves. The system

uses a simple web camera/mobile camera for gesture capturing and does not require high cost cameras. And the algorithm does not depend on skin tone of any person and hence the image processing is independent of the illumination.

2. LITERATURE SURVEY

Image processing can be described as every possible action performed on an image. This can be as simple as cropping an image, increasing contrast or scaling. Ever since digitalization of images came into the computer world, there was demand for object recognition. Object recognition is a classical problem in image processing. While humans can easily extract objects from an image, computers can't. For a computer an image is no more than a matrix of pixels. The artificial intelligence required for recognizing objects in this matrix, has to be created by a programmer. The object recognition approaches can be broadly classified into "Data-Glove based" and "Vision-based" [1]. Tracking bare hand and recognizing the hand gestures using low level features such as color, shape, or depth information [2] generally require uniform background, invariable illumination, a single person in the camera view, and/or a single large centered hand in the camera view. A lot of researchers initially used morphological operations to detect hand from image frames. The main drawback of this method lies in its huge computational complexity. To detect hand in a cluttered background, N.Petersen and D.Stricker used color information and histogram distribution model [3]. Some Local orientation histogram technique is [4] also used for static gesture recognition. These algorithms perform well in a controlled lighting condition, but fails in case of illumination changes, scaling and rotation. Chan and Ranganath used Fourier descriptors of binary hand blobs as feature vector to Radial Basis Function (RBF) classifier for pose classification and combined HMM classifiers for gesture classification [5]. Even though their system achieves good performance, it is not robust against multi variations during hand movement. To overcome the problem of multi variations like rotation, scaling, translation some popular techniques like SIFT [6], Haar-like features [7] with Adaboost classifiers [8] which is based on popular Viola Jones Algorithm [9]. The Haar-like feature[7] based and SIFT feature[10] based methods on Viola Jones algorithm are computationally expensive and time consuming. The solution is to implement the popular Viola Jones algorithm with LBP features which is very less computationally expensive.

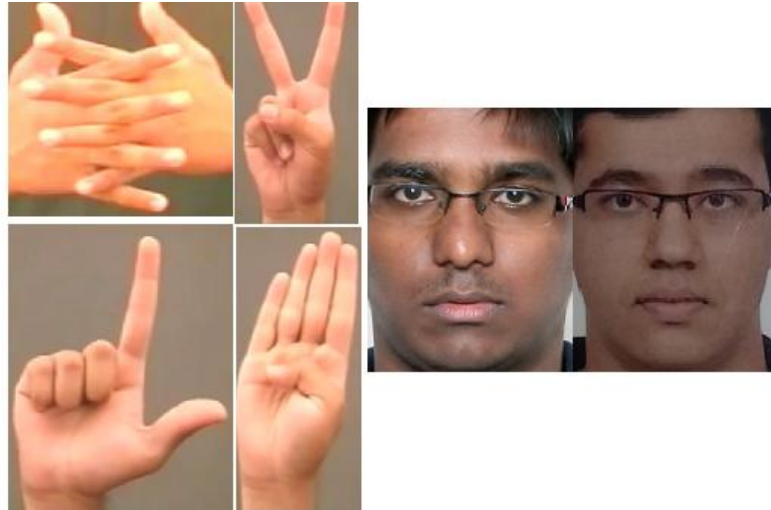


Fig 1: Hand and Face training images

3. TECHNICAL BACKGROUND

3.1 Viola-Jones Algorithm

Viola and Jones have proposed famous face detector architecture in 2004. The Viola-Jones face detector achieve 90% face detection rate and less than 50 false detection in 18,901,947 sub-windows scanned. (Viola and Jones, 2004) Due to its high accuracy on face detection, this approach is suitable to detect some other objects. Despite this success, it may not have same performance in gesture detection. The result is different from the method's performance on face detection. The reason is that the comparison of the training image between hand and face which shown in Fig 1, the hand training images contain much more background than face images. Because of the characteristic of face can be bounded by a rectangle box perfectly and no any background noise exists in the box. If the training image set contain much more noise, it is difficult to maintain high performance during the detection. Therefore, separate the background and the hand is required to achieve more accurate gesture detection.

3.2 The Adaptive Boosting Learning Algorithm

The Adaptive Boosting (Adaboost) learning algorithms can integrate the information of a category of objects. This algorithm originally used by Viola-Jones algorithm to train the sample set which involves cascade based classifier. It can combine the weak classifiers which cannot provide satisfactory result to become a strong classifier to get the better result. The Adaboost learning algorithm chooses the best weak classifier from a set of positive and negative images. After choosing the best weak classifier, the weights of the training images are adjusted by Adaboost algorithms. In this round the weights of classified training images are decreased and the unclassified images are increased. In the next round, the unclassified images will be more focused by the Adaboost and try to correctly classify the misclassified images. The whole procedures are finished until a predefined performance is satisfied. However, Ko-ChihWang (2007) [10] reported that the result of using AdaBoost with Viola Jones detector in hand detection is worse than face detection accuracy due to the structural problem of hand. And, he proposed using AdaBoost with SIFT Feature is more accurately. Therefore, it is necessary to apply AdaBoost

learning algorithm with suitable detector for increasing the detection accuracy. The LBP feature has been chosen to apply AdaBoost and hence show that with LBP feature detector, one can achieve fairly accurate hand detection is possible with computationally less expensive detectors like LBP.

3.3 Local binary patterns (LBP)

Local binary patterns (LBP) is a type of feature used for classification in computer vision. LBP is the particular case of the Texture Spectrum model proposed in 1990. LBP was first described in 1994. It has since been found to be a powerful feature for texture classification. The Local Binary Pattern (LBP), introduced by Ojala [13], serves for texture representation. The LBP is used across various computer vision fields (e.g. image synthesis, light normalization, face detection, face/expression recognition). In this work LBP descriptors are experimented on, to show its discriminative power for hand shape classification (i.e. capability to capture enough information in order to distinguish among hand shapes in the classification phase).

4. IMPLEMENTATION

For the gesture recognition, the Viola-Jones algorithm is used with LBP features to training the strong classifier. The following are the steps for training the classifier cascades[14]:

1. Training data preparation
2. Image segmentation
3. Creates sample
4. Training sample and use the strong cascade

4.1 Data preparation

First, in the data preparation, it is necessary to prepare vast original images used for Adaboost training. A large amount of original image is to improve the detection accuracy. It separated into two types, the positive and the negative images. The positive image is the detecting object image, for example, the hand detection needs many hand pictures as the positive image. Another is the negative images which haven't contained any object picture. In this paper, the focus will be on 'M' gesture detection. It is required to take positive images via a webcam and the background images from internet or a webcam.

4.2 Image Segmentation

The Second, in the image segmentation stage, it is important for reduce the background noise of the original image. Therefore, the pure background color is required during the image capture progress. Moreover, the lighting condition and background noise also will influence the quality of the training data such as a hand and the background color is not separated clearly due to the lighting condition. To tackle the problem, HSV model's threshold value is used to separate the background noise in the training images.

4.2 Create Samples

Third, in create sample stage; a description file is created for those positive images. After generating the description file such as positive.txt, this description file is used to create a file name positive.vec to store the sample file. This positive.vec files which stores the entire sample is for the sign gesture recognition. Therefore, it is possible to produce each classifier for each gesture.

4.4 Training sample and use the strong cascade

Fourth, it is time to create the classifier by using the LBP training in OpenCV. When using the LBP training command, a lot of negative images need to be set up, more background image, and the better result. After the process is finish, the LBP training creates an .xml file. This classifier is cascaded stronger classifier. At last, the stronger classifier is used to detection the hand gesture; it is the major issue for gesture recognition.

5. PERFORMANCE ANALYSIS

The sign M has been chosen for performance analysis. The major parameters to be considered during performance analysis are

1. Number of positive images
2. Number of negative images
3. Number of stage

The performance of the classifier for M gesture is tested by considering different number of stages, positive images and negative images.

First let us consider the dependence of Hit Rate on the parameters 2 and 3 while keeping the parameter 1 (Number of positive images constant). The numbers of positive images are taken in 3 different sets numbered 802,999 and 1263

It can be seen from Fig 2 that the hit rate increases with the number of stages. It is lowest at 4 stages and reaches a maximum of 52.3% at 12 stages and 2382 negative images before rolling off.

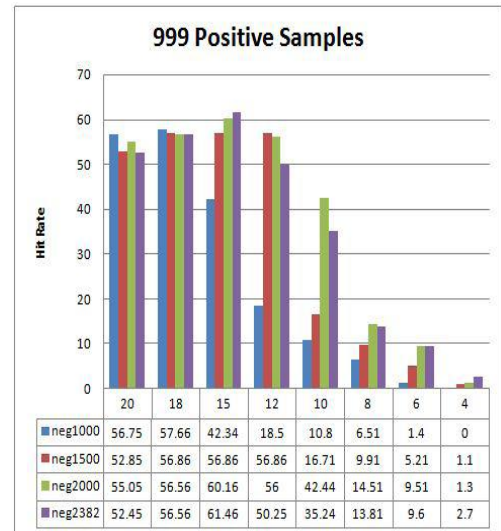


Fig 2: Hit Rate vs. No of Stages (802 Positive Samples)

In Fig 3, it can be seen that the highest hit rate is achieved at 61.46% for 15 stages and 2382 negative images before decreasing gradually.

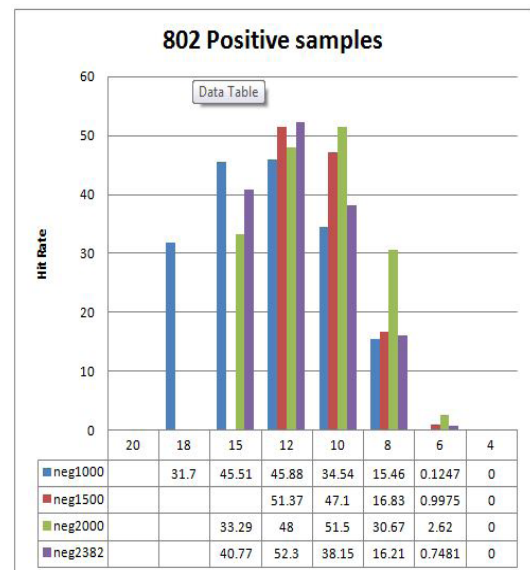


Fig 3: Hit Rate vs. No of Stages (999 Positive Samples)

By keeping the number of negative images constant and increasing the number of stages, no improvement in the situation can be seen. A rapid rise in hit rate can be seen when the number of stages is increased from 8 to 10 with the number of negative samples at 2000. The hit rate is still low at these two stages, but comparatively higher than with 802 positive samples. Also, the maximum hit rate is increased by 9.16% as a result of taking a higher number of positive samples

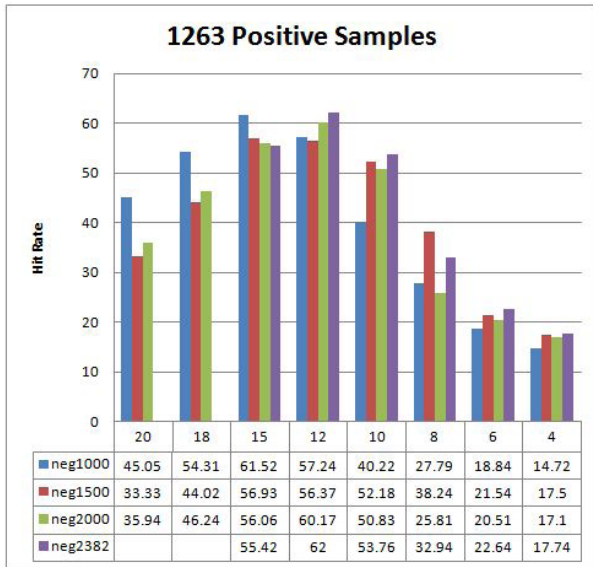


Fig 4: Hit Rate vs. No of Stages (1263 Positive Samples)

In Fig 4, with 1263 positive samples, the highest hit rate achieved is 62% at 12 stages and 2382 negative samples before rolling off. The increase in hit rate from 999 positive samples and 1263 positive samples is marginal, than with 802 and 999. The minimum Hit rate still continues to rise with the rise in positive images as can be ascertained with the above 3 graphs.

This implies that the Hit rate is directly proportional to the number of positive samples until a certain criterion is met, after which it goes on decreasing. By taking the highest hit rates from all 3 cases into account, it is safe to conclude that there exists a criterion to obtain an optimal Hit Rate. This is found to be when the positive and negative images are in 1:2 ratios. Any change in this ratio only serves to decrease the Hit Rate rather than increasing it.

Now, let us analyze the False Alarm Rate in the same manner

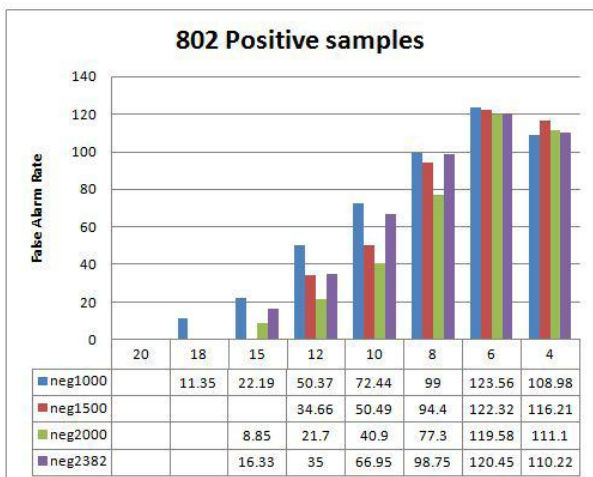


Fig 5: False Alarm Rate vs. No of Stages (802 Positive Samples)

From Fig 5, it can be seen that the false alarm rate is highest at 123.56% stages before it goes on reducing until it reaches a minimum of 11.35% at 18 stages

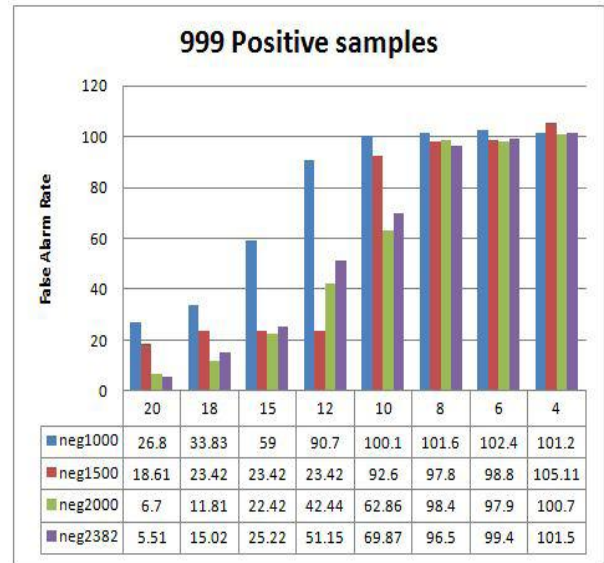


Fig 6: False Alarm Rate vs. No of Stages (999 Positive Samples)

Similarly, in Fig 6, it can be seen that the False Alarm Rate is highest at 4 stages before reaching a minimum at 20 stages. The maximum False Alarm Rate has decreased by 18.45%.

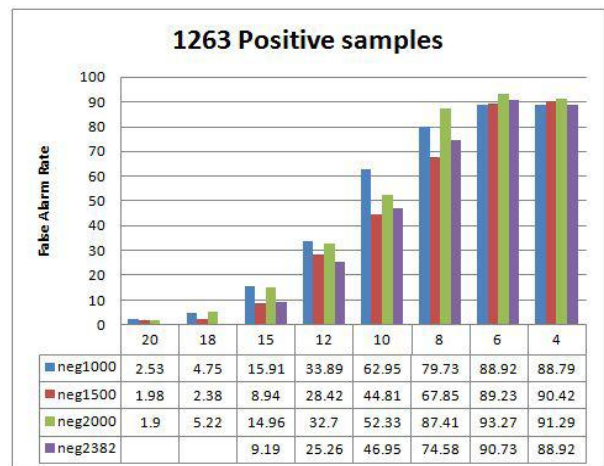


Fig 7: False Alarm Rate vs. No of Stages (1263 Positive Samples)

Finally, in Fig 7, it can be seen that the maximum False alarm rate is still decreased to 93.27% as compared to 123.56% at 802 positive samples. The false alarm rate is minimum at 1.9% and 20 stages from graphs Fig 5, Fig 6 and Fig 7; it can be concluded that the false alarm rate goes on decreasing with the number of positive images. This means that the false alarm rate is inversely proportional to the number of positive samples. Also, the best result is achieved when the ratio of positive and negative samples is in the ratio of 1:2

By considering the Hit rate and False Alarm Rate analysis for all the above cases, one can conclude that a strong classifier with a high hit rate and proportionally smaller False Alarm Rate can be achieved at 999 positive images, 2000 negative images and 15 stages. Any case other than this has comparatively worse performance. This scenario can be achieved with the ratio of positive and negative samples kept at 1:2

6. EXPERIMENT

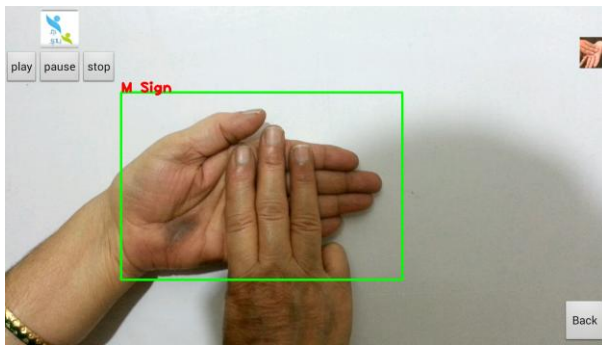


Fig 8: M-gesture (Ideal Background)

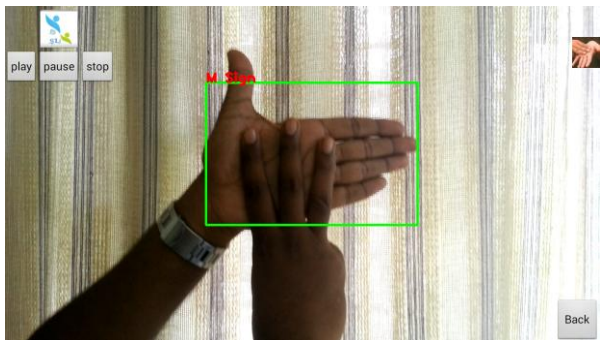


Fig 9: M-gesture (Ideal Background)

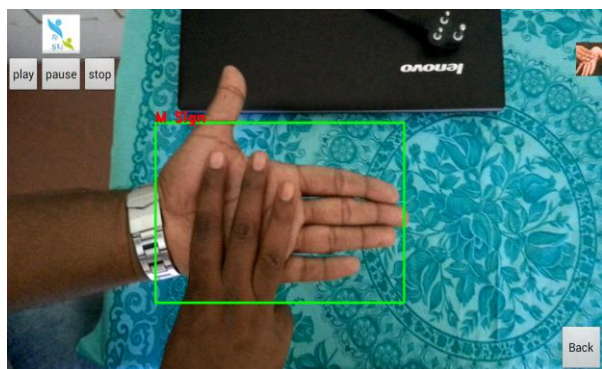


Fig 10: M-gesture (Complex Background)

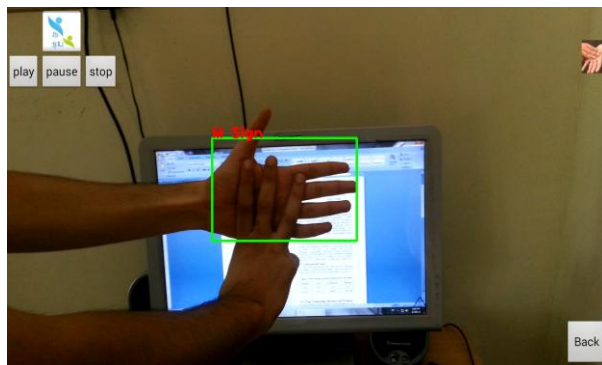


Fig 11: M-gesture (Complex Background)

An android application has been designed to demonstrate the implementation of Viola-Jones algorithm with LBP features for hand gesture recognition. The OpenCV libraries [11] have been used to develop the application and android app is

available in Google Play store [12] for free. During runtime, camera continuously captures the image of sign formation which will be processed and matched with the database.

Here, for database the .xml file generated after the cascade training will be used. Further, detected gesture will be marked with the rectangle and appropriate sign will be named. From Fig 8, 9 10 and 11, it can be concluded that M sign will be detected even in complex and noisy background. Our implemented algorithm is robust and real time recognition is possible.

7. CONCLUSION

From our performance analysis, it is concluded that the optimized classifier can be obtained with less number of positive samples provided by considering the suitable number of stages and 1:2 ratio of positive to negative images.

To demonstrate the concept an android application has been developed. From android experiment, it has been showcased that detection is possible even in complex backgrounds and in noisy condition. Android application implementation has been extended to 4 other sign letters i.e. A, B, V and W. It has been proved through this that sign recognition depends on mobile RAM capacity. With high RAM size more number of sign letters can be accommodated into single android activity. It has been analyzed that with modification to LBP features it is possible to extend the above implemented algorithm to more number of signs. Implemented algorithm can be further extended to general object recognition. Future applications include real time face recognition containing large databases for security purposes, enhanced gesture recognition implementation in TV and other electronic devices to allow a seamless experience

8. REFERENCES

- [1] R.S. Jadon G.R.S.Murthy. A review of vision based hand gestures recognition, international journal of information technology and knowledge emangement. pages Vol 2, pp. 405-410, July-December 2009.
- [2] J.Cheng H.Lu Y.Fang, K.Wang. A real time hand gesture recognition method. In IEEE, 2007.
- [3] Didier Stricker Nils Petersen. Fast hand detection using posture invariant constraints. In Springer-Verlag Berlin Heidelberg LNAI 5803, pages pp.106-113, 2009.
- [4] Hanning Zhou, D.J. Lin, and T.S. Huang. Static hand gesture recognition based on local orientation histogram feature distribution model. In Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on, pages 161-161, June 2004. doi: 10.1109/CVPR.2004.169.
- [5] Surendra Ranganath Chan Wah Ng. Real-time gesture recognition system and application. In Image and Vision Computing 20, pages 993-1007, 2002.
- [6] David G.Lowe. Distinctive image features from scale-invariant keypoints. In International Journal of Computer Vision 60(2), pages 91-110, 2004.
- [7] M.Petriu Q.Chen, N.Georganas. Real time vision based hand gesture recognition using haarlike features. In Instrumentation and Measurement Technology Conference { IMTC, Warsaw, Poland, pages 91-110, May 1-3, 2007.
- [8] R.Verschae H.Francke, J.Ruiz-del-solar. Real- time hand gesture detection and recognition using boosted

- classifiers and active learning. In Springer-Verlag Berlin Heidelberg LNCS 4872, pages 533-547.
- [9] J Michael Jones Paul Viola. In Second International Workshop on Statistical and Computational Theories of Vision- Modeling, Learning, Computing and Sampling, Vancouver, Canada.
- [12] Sign Language Interpreter.
<https://play.google.com/store/apps/details?id=com.opencv.analogicalnexus.facedetect>
- [13] T. Ojala, M. Pietikäinen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR
- [10] Chieh-Chih Wang and Ko-Chih Wang. Hand posture recognition using adaboost with sift for human robot interaction.
- [11] Gary Bradski and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc., 2008.
- International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 - 585.
- [14] Jin, Man Mau. "Real Time Hand detection and gesture recognition system." (2011). Retrieved from City University of Hong Kong, CityU Institutional Repository