Grenze International Journal of Engineering and Technology, Jan 2020

Indian Sign Recognition System using Adaptable Segmentation and Machine Learning

Rishabh Bhatt[1] and Dr. Vishal Bharti[2]
[1] Vellore Institute of Technology, Vellore, India
Email: rishabhbhatt009@gmail.com
[2] DIT University, Dehradun, India
Email: vishal.bharti@dituniversity.edu.in

***Abstract*—Deaf community all over the world uses sign language as a means of communication. Although it resolves the communication barrier among the community itself, the integration of acoustic and sign language is still matter of research. This paper proposes a dynamically adaptable image recognition module coupled with convolution neural network to recognize hand signs. The paper addresses the problem of isolation of hand from a robust environment and recognizing the hand sign. A series of transformation and segmentation is applied to the input, which in turns helps in feature extraction. Pre-processing of both training and input and is done to give a performance boost. Comparative study is drawn in the end to find the effect of pre-processing, of input and training data, on the classification speed.**

***Index Terms*— Hand Sign Recognition; Neural Network; Image Segmentation; Sign Language; Hand Tracking.**

I. INTRODUCTION

Effective communication is necessary for the betterment of establishment. The vocal words of people do matter and carries some weightage. For most of these instances acoustic language is being used by the general populous. Deaf community has found difficulty in communicating their thoughts, ideas and decisions properly. Introduction of sign language by various countries have aided the deaf community. Their inability to use vocal voices can be resolved either by write ups or a hearing aid device or by hiring a translator. Hearing aid devices and hiring a translator turn out to be expensive solutions whereas writing up can be more time consuming. Unlike spoken languages sign language does not have readily available application for translation. Due to which a gigantic communication barrier is created. Hence the researchers are trying to create a computer system as a translator.

The existing solutions were broadly classified into two categories: *sensor based* and *vision based*. Both approaches have the classification module in common however a sensor-based approach uses a glove fitted with accelerometers and other sensors whereas a vision-based approach uses a camera and is highly dependent on image processing. Although a sensor-based approach is accurate however the expensive hardware makes the technology not suitable for common people. [2] makes us aware of the limitations of sensor-based technology whereas [3] emphasizes on the importance of skin segmentation and sign recognition for pre-processing module.

Grenze ID: 01.GIJET.6.1.3

The performance of the latter depends highly on the dataset and mode of input. [21] tells us the effects of degraded dataset on performance of classifier. Thus, the inter dependency between two modules makes it necessary to have pre-processing of input and training data. Vision based approaches generally use a dataset of images of different users. Depending on the classifier used, various properties of images are studied to classify them. The recognition module has live video from the camera as a mode of input. The video is basically image frames therefore a convolution neural network is preferred over any other classifier since it has shown remarkable results. It finds its limitation in dynamically adapting according to spatiotemporal conditions and to isolate hand when overlapping with face during recognition. Recent developments in segmentation and morphological transformations of images have led to the birth of various skin-segmentation based methodologies.

In the following paper next sections tells us about the existing system solutions. Third Section tells us about the proposed system, histogram back projection is used to extract skin and morphology is used extract hidden features before being fed to classifier. Fourth Section tells us about classifier and its structure. The pre-processing is also applied to training data samples. We are expecting that this will provide a speedy and more accurate classification. Fifth section tells us about the accuracy and a comparative study is drawn to find an expected speed up. The last section summarizes the paper and concludes the results.

II. Related Works

The related work for the following process is divided into two sections:

A. Image Recognition

The authors of [1] proposes a method in which the input image is changed in HSV format. Hue and saturation value are tested since value represents brightness of image and does not have much effect in recognition. An input hand image is segmented using skin detection algorithm. Local binary histogram of the segmented image is calculated, which are grouped together to form one single vector, to find statistical property such as variance, standard deviation, kurtosis, entropy etc. of the image.

Paper [2] mentions the limitation of sensor-based technology. Sensor based technology highly relies on the environment and placement of sensors. Due to this property lots of people don't prefer gloves, accelerometers, EMG sensors and don't find it comfortable to wear it all the time. A method for Arabian sign language recognition, which involves hand segmentation and in-depth facial recognition using Kinect2 sensor provides and accuracy of 55% is achieved using random forest classification, doesn't rely much on gloves. Thresholding is use to identify hand and sign region. The system failed when the recognition system failed to properly capture the motion frames, hence proving how pivotal position recognition module holds.

[3] emphasize on the importance of skin segmentation and sign recognition for classification systems. They proposed system gives an image which is thresholded to identify skin color region. The input of various frames is compared to distinguish between static and movable objects. Paper proposes the usage of support vector machines infused with motion, color and position information of the image. Image is segmented using Kalman Filter. A is system proposed in [4] for sign recognition through video input. It states the importance of some but all spatial features of image in recognition. Hand signs are traced and segmented and input in the form of subunits is given to the system. Clustering on the spatiotemporal data is performed which is calculated through segmented subunits. This results in the formation of Codebook. System makes weak classifiers combining different features extracted from Codebook. Adaboost is used to transform the classifier into a strong classifier.

[5] provides a solution to extract hand features when they are overlapping face by using back propagation histogram and Kalman filter. Plotting of histogram and comparing it with previous observations to obtain pattern in properties has been adopted as a solution by many whereas usage of Kalman Filter to get the skin color object is most preferable.

B. Image Classification

Lots of classification techniques have been used by researchers to find out the optimum one which can be used for image classification. Classification image highly depends on the dataset and the computation power of the machine.

[1] and [3] used Support vector machine as a way to classify image. The single combined vector of statistical properties was fed in to the multi SVM to classify the image into seven categories. Elpetalgy at [2] mentions

a 55% accurate classifier for a data set of 150 made up on Random forest. The limitation proposed by them was somewhere in the recognition system.
Neural Networks have made the process of image classification easy. With the availability of computational resources and powers sources, one can achieve high level of success. [10] explains the function of pooling, convolution and fully connected layer in a neural network classifier. Grey scaled converted images were used. [4] tells us the advantage of using Linear Bayesian classifier. Less computation power is utilized by Bayesian classifier than neural network or back propagation algorithms. [21] tells us the effect of degraded dataset on performance of classifier. Since we aim to develop a robust system, thus incoming of noisy data for classification both for testing or training need a pre-processing. This pre-processing of the incoming input is highly dependent on environment conditions. Thus, in the following section we propose the use of k-means clustering which identifies major skin-color contours in the frames and then segment image again according to the lighter of two skin shades before being fed to the classifier.

III. METHODOLOGY

Computer vision-based approaches using image processing have found to be successful in providing promising results. The major issues bugging those types of solutions are the dynamic adaptability of the segmentation process to isolate skin-color contour irrespective of the environment, isolation of hand region when overlapping with face during the time of imitation and how to increase the speed of recognition in terms of frames per second.
Here with the use of histogram back projection makes the system dynamically adaptable. The addition of pre-processing is expected to increase accuracy of prediction and boost its speed.
For better understanding the Framework of the proposed system, we can divide the study in two different modules.

A. Image Processing Module

Image processing module is responsible for isolation of hand sign from the input frames. It uses a combination of unsupervised learning with image processing methods such as segmentation and morphological transformations. The introduction of this module makes the proposed system self-adaptive. Feature extraction of hand in robust environment, which can change dynamically, becomes a bottleneck because different spatiotemporal conditions have a huge effect on image properties. As segmentation is highly dependent on image properties, a slight change in environment conditions drastically have an impact on segmentation of image. Thus, by utilizing histogram of the current environment, a dynamic thresholding module is being proposed.
1. Calibration step and Face removal using Haar Cascade: Isolation of hand becomes a troublesome issue. Most of the systems have controlled environment or some hardware assisted measure to imitate the hand sign. To introduce dynamic adaptability to the system a calibration step is introduced. The utility of this step is to make system well aware of the surroundings and the object that needs to be identified. In the calibration step, whole is divided into smaller regions and a color-based histogram is plot and saved. As color-based histogram are highly dependent on surrounding, thus same objects will have different histogram in different environment. Once the system has the histogram of the object needed to be identified, it can easily segment all the other value pixels. As face and hand sign will fall into same histogram space, as both are of similar color frequency, removal of face becomes a task of utmost priority. The problem is solved by using Haar Cascade classifier. These are pre-trained models that can be used to identify the face in an image. The face here is recognized and then converted black to ease the isolation of hand.
2. Hand Sign Isolation using Histogram Back Projection: Once the system saves the histogram, it can be used to track and identify the object in the input. Histogram back projection is used to identify the hand region. The histogram is projected again on the input the pixel values and calculate in the probability of the pixels. region of interest is segmented giving us the corresponding hand.
3. Morphological Transformation: Segmented image has a lot of noise. Thus, a series of morphological transformations is performed for noise reduction. All images are converted to binary from HSV and then transformations are applied. Erosion is applied using 5X5 filter. And then dilation is applied using 5X5filter, thus creating an opening.

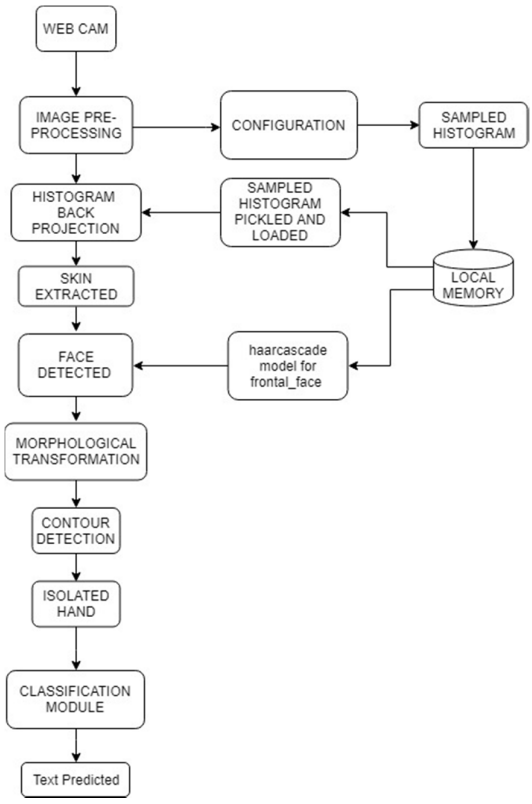

Figure 1 Flow Diagram of the system

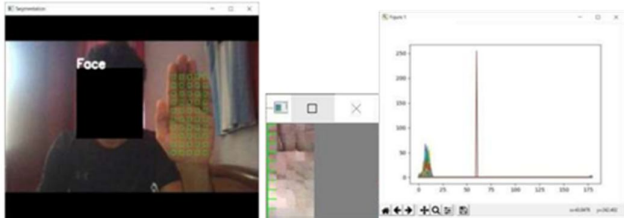

Figure 2 (A, B, C)- Show the configuration step
A – Configuration; B-Sampled Skin; Histogram Generated

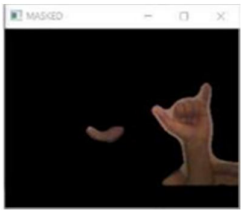

Figure 3 Segmented Hand

Figure 4 Morphological Transformation

B. Classification Module

Classification module builds the foundation of the system. With a total of 84,000 images 3000 images for each hand sign, which are 28 in number including space and no sign, are used at a ratio of 4:1 to train and test the classifier. While developing learning models using CNN, a hit and trial approach is followed since there is no proper benchmarks of how different entities of CNN impact one another. Once the model is prepared, time consumed in prediction reduces enormously.

1. Data Generation: CNN learns over images in the form of a vectors. Here pickle is used to store data. The data is divided into two smaller datasets: X and Y where former contains the extracted features and other contains the associated labels with it. The dictionary is formed and then transformed into a vector and saved. This step is known as pickling and helps us to reduce the time wasted in generating data again and again. The pickling of the data can be simply done hence making the job of data generation a onetime affair.

2. Training: Training of the model is highly parametric. Different configurations lead to different results. A CNN model yields better result when has hidden layer nodes more than nodes in the input layer and less than twice the number of nodes in the input layer. Here 4 layers in total are used including input and output layer. Various parameters with preferred configuration reasoning are mentioned below:

2. 1.Number of input nodes: The number of input nodes defines the size of the feature vector. Feature vector refers to the total dimension of features that are taken in consideration while learning. A dataset with lesser feature dimension to being with will provide better results than a dataset with higher feature dimension. Here the size of feature vector is 16. Number of input nodes (N_i) = 16

2. 2.Number of output nodes: Number of output nodes refers to the number of classes in which the input can be classified. When the number of classes is two it is called binary classification and if more than two it is referred as multi-class classification. Here the input frame can be categorized into 28 different classes. Therefore, Number of output nodes (N_o) = 28

2. 3.Number of hidden layers: Number of hidden layers is responsible for major learning and entropy changes during epochs. Higher the number of hidden layers, deeper the learning model will be. Generally neural networks with two hidden layers have to found working pretty well in the case of image classification.

Number of nodes in hidden layer:

$$(Ni) < (Nh) \leq 2(Ni)$$

Researchers have shown that there is no particular way of determining the number of nodes in hidden layer. Though having nodes twice the number of nodes in input layer does gives promising results. Here we have four layers in which one input (16 nodes), output (28 nodes) and two hidden layers (32 and 63 nodes) are present. A dense layer of 128 nodes is also attached before the output layer to combine the results. A dropout of 20% of values is done at random to avoid over fitting.

2. 4.Type of activation function: Activation functions refer to the mathematical equations responsible for progressive learning of the classifier. These functions are responsible to provide discrete output from the corresponding input. Hence the nodes are able to determine whether the data they are analyzing is useful or not. In the case of images, pixel value can be used to determine whether to consider the pixel in particular range or not. Here we have used two activation functions. ReLU which stands for Rectified Linear Unit is a non-linear activation function. This activation function is being used in input layer, hidden layers and the 128-node dense layer. ReLU helps to converge to result faster than other non-linear activation function, proving it to be more advantageous than others. SoftMax is the other activation function used in the output layer. It's ability and compatibility to handle multiple classes makes it the best alternative for output layer. It normalizes the output by diving it by the sum of their output, which leaves only one significant value.

$$r(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{for } i = 1, \ldots, K \qquad \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K \tag{1}$$ (2)

$$f(x) = \log(1 + e^x) \tag{3}$$

2. 5.Filter size: Filter size tells us about the kernel dimension. The kernel dimension refers to the amount of environment considered while analyzing certain object. Here filter of three sizes is used: 2 X 2, 3 X 3

and 5 X 5. Different sizes of filter help in extracting different features. Filter size of 2 X 2 helps us to identify edges, whereas 3 X 3 helps us to identify corners.

2. 6. *Optimizer used:* Optimizer function are the function that aid in reaching the goal faster. Here the objective function is to reduce the batch loss and increase the batch accuracy in the training module. Two types of optimizers are available to address the following issues which are: first derivative and second derivative optimizers. Second derivative optimizers are really costly in terms of computation though they do provide better result in terms of converging. Thus, efficiency of second derivative optimizer for sparse dataset is much higher than efficiency of first derivative optimizer. But since image recognition is highly similar pattern recognition through pixel values, data is adequate. Hence here we have used SGD (stochastic gradient descent). The usage of SGD provides us with less usage of computational power and saves cost.

$$Q(w) = \frac{1}{n}\sum_{i=1}^{n} Q_i(w), \tag{4}$$

$$w := w - \eta \nabla Q(w) = w - \eta \sum_{i=1}^{n} \nabla Q_i(w)/n, \tag{5}$$

2. 7. *Type of Loss Function:* Entropy functions are responsible to learn the slightest of changes that occur up on changing certain values of the features. To address the classification type problem for images two type of cross-entropy functions are available which are binary-cross entropy and categorical cross-entropy. The former works better when the output is two classes, the latter works well when model is a multi-class classifier. Here as the output classes are more than hence categorical cross-entropy is being used. Categorical cross-entropy when coupled with SoftMax, called SoftMax activation, is used to train model to calculate probability over a number of output classes. The sum of all probabilities is equal to one.

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} \mathbf{1}_{y_i \in C_c} \log p_{model}[y_i \in C_c] \tag{6}$$

2. 8. *Learning rate:* Learning rate defines the slope of error function. Gradient descent helps to identify the global minima of the function but a higher learning rate can result in increase of loss function rather than decrease. Here a learning rate of 0.012 is used.

2. 9. *Max pooling:* Max Pooling helps us to reduce the dimensionality of major factors that are needed to be studied. It creates an abstract of possible outcomes out of which only most favorable ones are chosen and further analyzed. Strides in max pooling steps help us to jump matrix during the time of pooling. Thus it helps in reducing time in repeatedly analyzing the pixels. Here pooling is done by 2, 3 and 5 sized filters in input, first hidden layer and second hidden layer respectively. The same number of strides is used as filter size.

2. 10. *Epochs:* Epochs refers to the completion of one forward and backward pass in the classifier. Number of epochs is highly dependent of the size of dataset and the expected outcome of accuracy. Here 7 epochs are used. With higher number of batch size, epochs are generally less to avoid over fitting.

2. 11. *Batch size:* Batch size refers to the number of input fed into the classifier in one instant.

2. 12. *Testing:* A total of 600 images are used to test the classifier. An accuracy of more than 80% is achieved. A dropout of random 20% of values is done to avoid over fitting. Up on testing 10 images of three classes B, C and Y, 8 for B, 10 for C and 5 for Y were the number of right classifications.

2. 13. *Prediction:* Once the model is trained it is saved. The saved model is then imported to help in prediction. Here the input frame is fed through pre-installed webcam on the system. The captured frame is converted in grayscale and resized to image size of 200 X 200. The classifier is used to predict the class of the upcoming frame. The segmentation module helps us to isolate hand sign from other skin color-based regions.

IV. Results and Discussion

A. Performance of Model

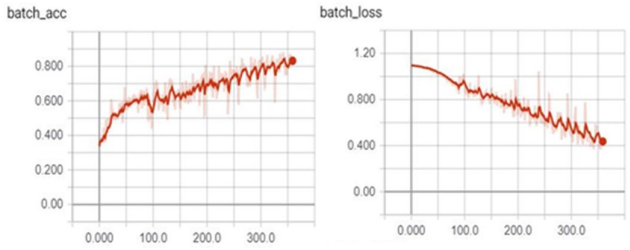

Figure 5 (A, B): A- Accuracy vs Iteration;
B- Accuracy vs Iteration

A graphical representation of varying batch accuracy and loss with respect to epochs. Smoothening of 80 per cent is applied. We can see that as the number of epochs rising, there is an increase in accuracy and a decrease in loss.

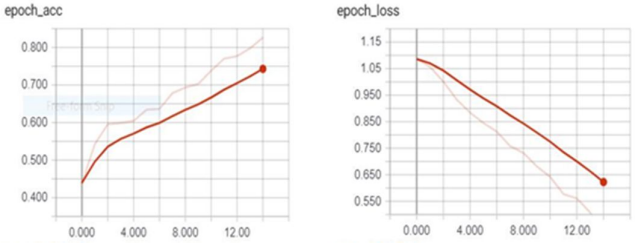

Figure 6 (A, B): A- Accuracy vs No. of epoch;
B- Accuracy vs no. of Epoch

A graphical representation of varying accuracy and loss with respect to epochs. Smoothening of 80 per cent is applied. We can see that as the number of epochs rising, there is an increase in accuracy and a decrease in loss. It remains constant for the last two epochs.

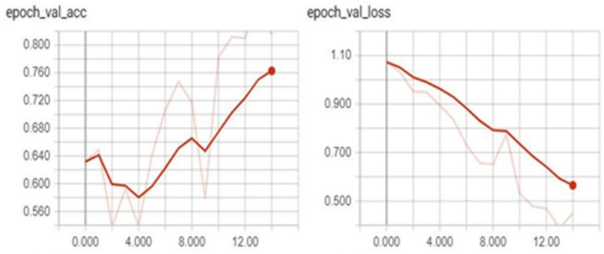

Figure 7 (A, B): A- Validation accuracy vs No. of epoch;
A- Validation accuracy vs No. of epoch

A graphical representation of varying value accuracy and loss with respect to epochs. Smoothening of 80 per cent is applied. We can see that as the number of epochs rising, there is an increase in accuracy and a decrease in loss. It remains constant for the last two epochs.

B. Snapshots

To test the validity of the system we took a set of images, screenshots from the live feed and ran it through the classification algorithm with and without pre-processing. The images without preprocessing showed a classification accuracy of 23% compared to 79% accuracy showed by the same images after pre-processing. The results show the gigantic difference in classification accuracy and the speed

1. Without pre-processing

Figure 8 College of screenshots of the live webcam feed

Figure 9 Classification results

2. With pre-processing

Figure 10 The segmented images after pre-processing

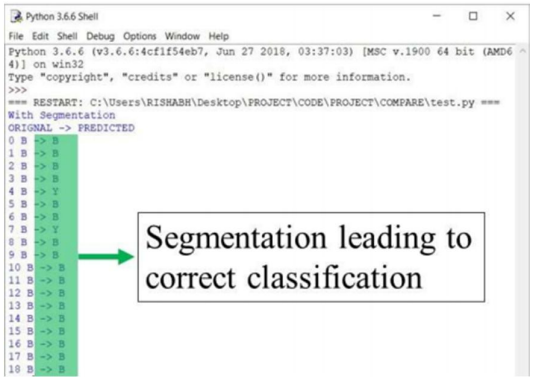

Figure 11 Classification Results

V. Conclusion

We were successfully able to isolate hand from robust environment conditions. The recognition module has an accuracy of over 80 per cent. The introduction of calibration step provides dynamic nature to our software solution and hence overcomes the limitation of
existing lab-controlled projects. The system does carry and initial cost of computation and training, but it's only for one time.

VI. Future Scope

The project is a raw implementation of recognition system with dynamic nature. Even though the system holds the ability to be adapt to new environments there are few areas where work can be done. The segmentation and isolation of hand when it hovers over the face is still not accurate. Moreover, the system only recognizes static hand signs and can be modified to recognize hand movements for hand signs that involve movement. The solution can be enhanced performance wise by providing better hardware support in terms of more computation power in training of the classifier and better capturing devices. The system can also be developed into an application with necessary advancements. With an addition of a Natural Language Processing these predictions can be processed and converted into words and sentences which can then be converted into speech. Thus, forming a fully functional sign language to speech translator which would work in real time.

References

[1] S. M. Saqlain Shah, H. Abbas Naqvi, J. I. Khan, M. Ramzan, Zulqarnain and H. U. Khan, "Shape Based Pakistan Sign Language Categorization Using Statistical Features and Support Vector Machines," in IEEE Access, vol. 6, pp. 59242-59252, 2018. doi: 10.1109/ACCESS.2018.2872670
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8480101&isnumber=8274985

[2] Elpeltagy, Marwa; Abdelwahab, Moataz; Hussein, Mohamed E.; Shoukry, Amin; Shoala, Asmaa; Galal, Moustafa: 'Multi-modality-based Arabic sign language recognition', IET Computer Vision, 2018, 12, (7), p. 1031-1039, DOI: 10.1049/iet-cvi.2017.0598 IET Digital Library,
URL:https://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2017.0598

[3] K. Imagawa, Shan Lu and S. Igi, "Color-based hands tracking system for sign language recognition," Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, 1998, pp. 462-467. doi: 10.1109/AFGR.1998.670991
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=670991&isnumber=14786

[4] Tanibata, Nobuhiko, Nobutaka Shimada and Yoshiaki Shirai. "Extraction of Hand Features for Recognition of Sign Language Words." (2002).

[5] Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, Peter Presti, "American Sign Language Recognition with the Kinect" ICMI '11 Proceedings of the 13th international conference on multimodal interfaces Pages 279-286 Alicante, Spain — November 14 - 18, 2011 ACM New York, NY, USA
ISBN: 978-1-4503-0641-6 doi: 10.1145/2070481.2070532

[6] J. -. Chu, I. Moon and M. -. Mun, "A Real-Time EMG Pattern Recognition System Based on Linear-Nonlinear Feature Projection for a Multifunction Myoelectric Hand," in IEEE Transactions on Biomedical Engineering, vol. 53, no. 11, pp. 2232-2239, Nov. 2006. doi: 10.1109/TBME.2006.883695
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1710164&isnumber=36082

[7] C. Chuan, E. Regina and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," 2014 13th International Conference on Machine Learning and Applications, Detroit, MI, 2014, pp. 541-544.doi: 10.1109/ICMLA.2014.110
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7033173&isnumber=7033074

[8] X. Chen, X. Zhang, Z. Zhao, J. Yang, V. Lantz and K. Wang, "Multiple Hand Gesture Recognition Based on Surface EMG Signal," 2007 1st International Conference on Bioinformatics and Biomedical Engineering, Wuhan, 2007, pp. 506-509.doi: 10.1109/ICBBE.2007.133
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4272617&isnumber=4272485

[9] H. Brashear, T. Starner, P. Lukowicz and H. Junker, "Using multiple sensors for mobile sign language recognition," Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings., White Plains, NY, USA, 2003, pp. 45-52.doi: 10.1109/ISWC.2003.1241392
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1241392&isnumber=27832

[10] J. Wu, L. Sun and R. Jafari, "A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors," in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 5, pp. 1281-1290, Sept. 2016.doi: 10.1109/JBHI.2016.2598302
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7552525&isnumber=7558265

[11] C. Chuan, E. Regina and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," 2014 13th International Conference on Machine Learning and Applications, Detroit, MI, 2014, pp. 541-544. doi: 10.1109/ICMLA.2014.110
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7033173&isnumber=7033074
[12] Yuntao Cui and Juyang Weng “Appearance-Based Hand Sign Recognition from Intensity Image Sequences” URL:https://doi.org/10.1006/cviu.2000.0837
[13] A. Barla, F. Odone and A. Verri, "Histogram intersection kernel for image classification," Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 2003, pp. III-513.doi: 10.1109/ICIP.2003.1247294
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1247294&isnumber=27938
[14] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang and Y. Gong, "Locality-constrained Linear Coding for image classification," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 3360-3367. doi: 10.1109/CVPR.2010.5540018
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5540018&isnumber=5539770
[15] J. Oh et al., "Avatar-Based Sign Language Interpretations for Weather Forecast and other TV Programs," in SMPTE Motion Imaging Journal, vol. 126, no. 1, pp. 57-62, Jan.-Feb. 2017.doi: 10.5594/JMI.2016.2632278 URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7847511&isnumber=7847455
[16] P. V. V. Kishore, D. A. Kumar, A. S. C. S. Sastry and E. K. Kumar, "Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition," in IEEE Sensors Journal, vol. 18, no. 8, pp. 3327-3337, 15 April15, 2018. doi: 10.1109/JSEN.2018.2810449
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8304621&isnumber=8322357
[17] R. K. Shangeetha., V. Valliammai. and S. Padmavathi., "Computer vision based approach for Indian Sign Language character recognition," 2012 International Conference on Machine Vision and Image Processing (MVIP), Taipei, 2012, pp. 181-184.doi: 10.1109/MVIP.2012.6428790
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6428790&isnumber=6428740
[18] X. Yang, X. Chen, X. Cao, S. Wei and X. Zhang, "Chinese Sign Language Recognition Based on an Optimized Tree-Structure Framework," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 4, pp. 994-1004, July 2017.doi: 10.1109/JBHI.2016.2560907
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7464244&isnumber=7963898
[19] B. K. Chakraborty, D. Sarma, M. K. Bhuyan and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human–computer interaction," in IET Computer Vision, vol. 12, no. 1, pp. 3-15, 2 2018. doi: 10.1049/iet-cvi.2017.0052
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8268688&isnumber=8268682
[20] O. Boiman, E. Shechtman and M. Irani, "In defense of Nearest-Neighbor based image classification," 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-8. doi: 10.1109/CVPR.2008.4587598
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4587598&isnumber=4587335