Hand Gesture Recognition for Deaf and Dumb using MATLAB

Yuvraj Singh Thakur,
UG Student, KJSIEIT, Mumbai
University of Mumbai, India.
yuvraj.t @somaiya.edu

Dhavit Vaghela,
UG Student, KJSIEIT, Mumbai
University of Mumbai, India.
dhavit.v @somaiya.edu

Satish Yadav
UG Student, KJSIEIT, Mumbai.
University of Mumbai, India.
satish.yadav @somaiya.edu

Shreyas Vaidya
UG Student, KJSIEIT, Mumbai
University of Mumbai, India
shreyas.v @somaiya.edu

Sejal Shah
Faculty, KJSIEIT, Mumbai
University of Mumbai, India.
sshah@somaiya.edu

***Abstract*— With the advent of image processing and computer vision techniques, there has been huge research going on in the field of biometric classification and identification. On combining these techniques with machine learning, not only efficient but also smart applications can be developed. Biometric refers to measurement of various human body parts such as hands, fingers, face, etc. to distinctive features such as fingerprints, iris scan, etc. A sub branch of this field is hand gesture recognition which deals with classification of hand movements and signs. This paper shows how to classify those hand gestures using Computer Vision and Machine Learning algorithms and convert them into text and voice to facilitate communication between the hearing deficient and normal people.**

Keywords: KNN, Gray scaling, GUI, HSI, SURF,

INTRODUCTION

The field of computer vision is growing day by day with constant research in the fields such as object detection[1], face detection[2], biometric authentication[3][4] and much more. Computer vision combines the human visual perception with speed of computing to draw accurate results from an image or a video feed. Machine learning helps in finding patterns in each set of training data and provide accurate results for a test data by learning from those patterns. Combining this characteristic of computer vision and machine learning can lead to smart, efficient and reliable image classification systems. Image classification is a process of classifying/finding a match for an input image matching its features with the features of previous training sets.As the problem uses to computer vision the first step is to capture an image using a camera. While traditionally low resolution cameras are used for image capturing, Bay, Herbert [5] approach uses Kinect. Kinect senses the depth at which the hand is located and thus helps in proper segmentation of and from the background. While webcam do not help in segmenting the image but the most important property they possess is, that they are cheap while Kinect are really expensive. Various image segmentation approaches such as blob detection, background subtraction etc can be used. Next step is to define various feature descriptors various feature descriptors based on edge information or using feature description methods such Harris methods [11] can be used. Mikolajczyk and Schmid[12] show comparison of such local descriptors For learning from these features, there are various methods such as Neural Networks, Machine learning, template matching etc.

SYSTEM DESIGN

The overall design and workflow process is divided into the block diagram shown in fig1.

A. Image Acquisition

In Image Acquisition, the test image of hand sign is captured using a webcam. A webcam is a digital video camera used for basic image and video capturing and functions. Other methods for image acquisition include using an depth sensing camera which is used to segments out the object with least depth[5].

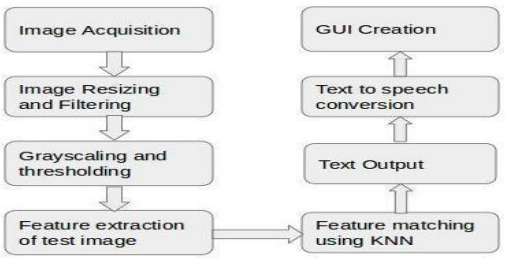

Fig.1 System Design Steps

B. Grayscaling and Thresholding

The test image taken is converted into a grayscale image and then thresholded to differentiate the hand sign from the background. For thresholding the image we are use Otsu's method[5]. Thresholding is used for segmenting out the hand from the background. Thresholding can only be used for applications where there is uniform background. If a non-uniform background is used, then other methods should be implemented such as skin colour detection by converting the image into YCbCr[9][10] format. From the given two the latter is most preferred. The reason behind this is that while for different colours of skin Y component varies but the Cb and Cr component show the least variation.

C. Image Resizing and Filtering

For faster computation of the hand signs, image resizing is done and the noise present in the image is removed by applying median filters to the image. Median filtering[6] is used for removing salt and pepper noise from any grayscale image. The median filter maintains the sharpness of the image while removing its noise[7].

D. Feature Extraction from Test Image

Feature extraction is a method of finding various distinctive features in the image. Here first the SURF[1] algorithm is applied on the image, this algorithm provides us with various feature descriptive features. The SURF algorithm is scale and orientation independent. SURF identifies the features in a image which show least effect of scale and orientation using Laplacian of Gaussian and 2x2 Hessian matrix on the input image.

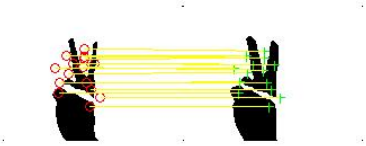

Fig.2 SURF features descriptors

E. Feature Matching Using KNN

KNN is one of the most used machine learning algorithm in the field of Statistical pattern recognition. The working of KNN is quite simple, the feature vectors of the training set are first trained to produce a classifier. Once a classifier has been produced then the feature vectors from a test image are extracted and given to the classifier. The classifier looks for K nearest feature vectors for the given test feature vectors. This is done by finding the distance between the test and trained feature vectors. Various distance functions[13] such as Hamming distance, Euclidean distance, city block distance can be used to find the nearest neighbouring feature vectors.

Consider the test image in fig(2), the image has two families of features: squares and triangles. Their houses are shown in the feature space. Consider a new feature creates a house in the feature space shown in the black circle. this house should be assigned to one of this feature family i.e. square/triangle. This process is called Classification. From the image, it is very easy to determine that square is the nearest neighbour to the new feature. But if many triangles are near to it, then it is not sufficient to find the nearest neighbour of the image. Instead we check some K nearest families of the image and assign the majority in them to the new feature in the space. In the image, supposes K = 5, which means 5 nearest neighbours need to be checked. He has three triangles and two squares, so now the new feature should be assigned to the triangles family. This is the method to calculate the k nearest neighbours. The value always changes with the change in the value of k. A database of the hand gestures representing numbers 0 to 9 was prepared as training examples.

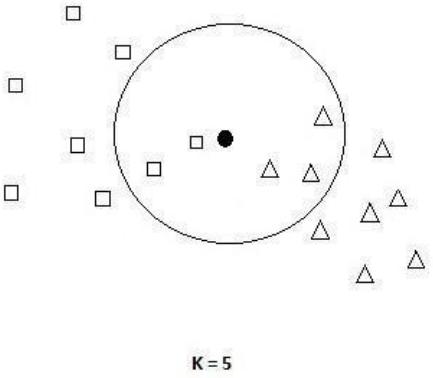

Fig.3 KNN

visualization *F. GUI creation*

A GUI (Graphical user interface) is created using MATLAB GUIDE app to make the project user friendly. GUI contains an option to convert the detected symbol from text to speech. For Text to speech we will be using text2speech - tts file from the MATLAB file exchange[15].

RESULT

The classifier is trained using the *bagoffeatures()* function of MATLAB for feature dimension independent classification. Bag of features converts the features into 500 visual words based on their occurrence in the images. Below is the confusion matrix for the above trained classifier.

	PREDICTED									
KNOWN	0	1	2	3	4	5	6	7	8	9
0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.94	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
2	0.06	0.00	0.82	0.06	0.06	0.00	0.00	0.00	0.00	0.00
3	0.06	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.00
4	0.06	0.00	0.00	0.12	0.71	0.00	0.12	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.94	0.06	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
7	0.00	0.00	0.06	0.00	0.06	0.00	0.00	0.88	0.00	0.00
8	0.06	0.06	0.00	0.00	0.00	0.00	0.06	0.00	0.76	0.06
9	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.88

* Average Accuracy is 0.89.

Fig.4 Confusion Matrix

While the confusion matrix indicates that efficiency of 89% was achieved after running the classifier on images captured through webcam while testing indicate the efficiency of around 72.4%. Below are the images indicationg various hand signs.

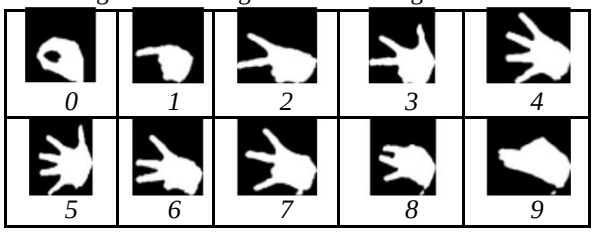

Fig.5 Hand Signs from 1 to 10

Fig.6 Output of classifier for above image

IV. CONCLUSION

An hand gesture classification method was developed using SURF feature description method. The classifier gave an efficiency of about 72.4% and was run on a number of text images. The downside of the method is the local features which depend on overlapping fingers could not be identified because the edges of overlapping figures gets filled due to image thresholding.

REFERENCES

Bay, Herbert, et al. "Speeded-up robust features (SURF)." *Computer vision and image understanding* 110.3 (2008): 346-359.

Kim, Donghoon, and Rozenn Dahyot. "Face components detection using SURF descriptors and SVMs." In *Machine Vision and Image Processing Conference, 2008. IMVIP'08. International*, pp. 51-56. IEEE, 2008.

Snelick, Robert, Umut Uludag, Alan Mink, Mike Indovina, and Anil Jain. "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems." *IEEE transactions on pattern analysis and machine intelligence* 27, no. 3 (2005): 450-455.

H a n , Chin-Chuan, Hsu-Liang Cheng, Chih-Lung Lin, and Kuo-Chin Fan. "Personal authentication using palm-print features." *Pattern recognition* 36, no. 2 (2003): 371-381.

Li, Yi. "Hand gesture recognition using Kinect." In *Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on*, pp. 196-199. IEEE, 2012.

Wang, Zhou, and David Zhang. "Progressive switching median filter for the removal of impulse noise from highly corrupted images." *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 46, no. (1999): 78-80.

[book] Sicuranza, Giovanni. *Nonlinear image processing*. Elsevier, 2000.

Otsu, Nobuyuki. "A threshold selection method from gray- level histograms." *IEEE transactions on systems, man, and cybernetics* 9, no. 1 (1979): 62-66.

Pradhan, Ashis, Shubham Kumar, Dependra Dhakal, and Bishal Pradhan. "Implementation of PCA for Recognition of Hand Gesture Representing Alphabets." *International Journal*6, no. 3 (2016)

Huang, C.L., Jeng, S.H., 2001. A model-based hand gesture recognition system. Machine Vision and Applications (12), 243– 258.

Harris, Chris, and Mike Stephens. "A combined corner and edge detector." In *Alvey vision conference*, vol. 15, no. 50, pp. 10-5244. 1988.

Mikolajczyk, Krystian, and Cordelia Schmid. "A performance evaluation of local descriptors." *IEEE transactions on pattern analysis and machine intelligence* 27, no. 10 (2005): 1615-1630.

Phung, Son Lam, Abdesselam Bouzerdoum, and Douglas Chai. "A novel skin color model in ycbcr color space and its application to human face detection." In *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, pp. I-I. IEEE, 2002.

Dardas, Nasser H., and Nicolas D. Georganas. "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques." *IEEE Transactions on Instrumentation and measurement* 60, no. 11 (2011): 3592-3607.

Fukunaga, Keinosuke. "Statistical pattern recognition." In *Handbook Of Pattern Recognition And Computer Vision*, pp. 33-60. 1999.

Zhang, Hao, Alexander C. Berg, Michael Maire, and Jitendra Malik. "SVM-KNN: Discriminative nearest
neighbor classification for visual category recognition." In *Computer Vision and Pattern
Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, pp. 2126-2136. IEEE, 200

Walters-Williams, Janett, and Yan Li. "Comparative study of distance functions for nearest neighbors." In *Advanced Techniques in Computing Sciences and Software Engineering*, pp. 79-84. Springer, Dordrecht, 2010.

Alfanindya, Alexandra, Noramiza Hashim, and Chikannan Eswaran. "Content Based Image Retrieval and Classification using speeded-up robust features
(SURF) and grouped bag-of- visual-words (GBoVW)." In *Technology, Informatics, Management, Engineering, and Environment (TIME-E), 2013 International Conference on*, pp. 77-82. IEEE, 2013.

links for TTS download
Microsoft's TTS Namespace
http://msdn.microsoft.com/en-us/library/system.speech.synthesis.ttsengine(v=vs.85).aspx
Microsoft's Synthesizer Class
http://msdn.microsoft.com/en-us/library/system.speech.synthesis.speechsynthesizer(v=vs.85).aspx
System Requrements:.NET or Microsoft SAPI (Speech SDK 5.1, Visual C++ 2005 or 2008) .