International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-3, Issue-5, June 2014

Artificial Neural Network Based Visual Recognition System using Dwt for Hearically Impaired Person

Y. M. Gaikwad, K. N. Pawar

Abstract:-Generally image processing is done to process an image for different application. There is variety of transform base feature extraction method. Visual recognition system or lip reading method is important generally in noisy condition. The new modality in image processing area is gives you dictation of voice. The discrete cosine transforms (DCT) and discrete wavelet transform (DWT) are techniques for converting a signal into elementary frequency components. These are widely used in image compression. Here we develop some functions to compute the DWT and to compress images. These functions illustrate the power of Mathematic in the prototyping of image processing algorithms. The rapid growth of digital imaging applications, including desktop publishing, multimedia, teleconferencing, and high-definition television (HDTV) has increased the need for effective and standardized image compression techniques.

Index terms:-ANN, DCT, DWT, HMM.

I. INTRODUCTION

There are generally two types of models.

1. Artificial neural network.
2. Hidden morkov model.

Proposed system of visual recognition consist of video signal as input signal .The flow of system can be explained with the help of following diagram. Figure 1 shows the different building blocks of recognition system .Video quality depends on the resolution and number of frames per second used to represent that video. Generally 25 frames per second video are considered as original video or liable video for operation purpose. Section 1 describes about video and frames section 2 describes about feature extraction methods section 3 describes about classifier and section 4 describes how to get recognition at the output.

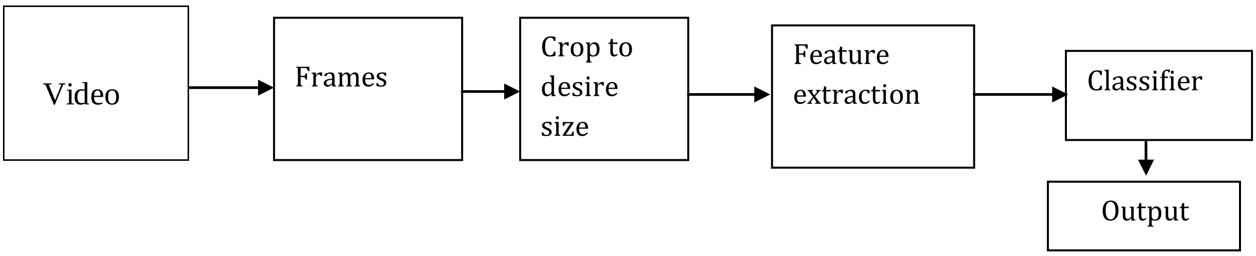

Fig. 1 General Visual Recognition System

II. VIDEO AND FRAMES

The facial expression of speaker is the main tool of recognition of speech hearically impaired aside cannot do this. we can develop a system to that person who can make interface with computer. If the audio signal is not clear or ideal the visual speech can become helpful tool in comprehending speech. In case of less signal to noise ratio there is large visual information. In 1954 sum by and Pollack Presented that bimodal video can be detected at very low signal to noise ratio (-30db. (1)

The number of frames created from the video, depend on frequency. Generally for 1 HZ frequency there are 95 frames.

Manuscript Received on June 2014.

Y. M. Gaikwad, M.E. Appear, Department of Electronics Engineering, S.S.V.P.S. B.S.D. C.O.E, Dhule, India.

K. N. Pawar, Head, Department of Electronics Engineering, S.S.V.P.S. B.S.D. C.O.E, Dhule, India.

III. FEATURE EXTRACTION

Lip region extraction can be done by two ways.

1) Image Transform based
2) Geometric based.

1) Image Transform Based

There are five transform which used mostly. Where as in image transform the redundant information is removed by converting each video frame in to lower dimensional space. (2)

1] DCT (Discrete Cosine Transform)
2] DWT (Discrete Wavelet Transform)
3] PCA (Principle Component Analysis)
4] LDA (Linear Discriminant Analysis)
5] FDCT (fast Discrete Curvelet Transform) (3)

2) Geometric Based

Geometric based parameterization considers geometry of face region. According to geometry of face region, it extracts lip region. It assigns some reference points on face region. By distance formula it extracts lip region. Example of this type of approaches is Active contours (often referred

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

to as snakes), deformable templates, and active shape models. Petajans original system is geometric based feature extraction method. [4] Where as [5] potamianos shows that image transform method is very good in noisy condition.

Result shows that image transform method is good. Out of above mentioned transform FDCT is non linear PCA and LDA are linear. In most Digital Signal Processing (DSP) applications, the frequency content of the signal is very important. The Fourier Transform is probably the most popular transform used to obtain the frequency spectrum of a signal. But the Fourier Transform is only suitable for stationary signals, i.e., signals whose frequency content does not change with time. The Fourier Transform, while it tells how much of each frequency exists in the signal, it does not tell at which time these frequency components occur. Over the past several years, the wavelet transform has gained widespread acceptance in signal processing in general and in image compression research in particular. In applications such as still image compression, discrete wavelets transform (DWT) based schemes have outperformed other coding schemes like the ones based on DCT. Since there is no need to divide the input image into non-overlapping 2-D blocks and its basis functions have variable length, wavelet-coding schemes at higher compression ratios avoid blocking artifacts. Because of their inherent multi-resolution nature, wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important. Recently the JPEG committee has released its new image coding standard, JPEG-2000, which has been based upon DWT.

1) *DWT Features for Mouth Region*

A DWT is a wavelet transform for which the wavelets are discretely sampled. The DWT of a signal is calculated by passing it through a series of low and high pass filters to obtain four sub bands viz., one approximation band and three detailed bands belonging to low frequency and high frequency components respectively. The four sub bands of DWT such as approximation band, horizontal band, vertical band and diagonal bands. The significant information of lip is present in the approximation band compared to other three high frequency component bands. The Discrete Wavelet Transform (DWT), which is based on sub-band coding, is found to yield a fast computation of Wavelet Transform. It is easy to implement and reduces the computation time and resources required. Wavelet transform decomposes signal into set of basis functions. These basic functions are called as wavelets. Wavelets are obtained by single prototype wavelet y(t) called mother wavelet.

$$\varphi a, b(t) = \frac{1}{\sqrt{a}}\varphi(\frac{t-b}{a}) \qquad (4)$$

Where a is scaling parameter and b is shifting parameter. The variable 'a' (inverse of frequency) reflects the scale (width) of a particular basis function such that its large value gives low frequencies and small value gives high frequencies. The variable 'b' specifies its translation along x-axis in time. The term $1/\sqrt{a}$ is used for normalization. The wavelet transform concentrates the energy of the image signals into a small number of wavelet coefficients. It has good time-frequency localization property. The fundamental idea behind wavelets is to analyse signal according to scale. It was developed as an alternative to the short time Fourier to overcome problems related to its frequency and time resolution properties. The advantage of DWT over DFT and DCT is that DWT performs a multi-resolution analysis of signal with localization in both time and frequency. Also, functions with discontinuities and with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sine-cosine basis vectors to achieve a comparable approximation. The properties of Wavelet Transform allow it to be successfully applied to non-stationary signals for analysis and processing, e.g., speech and image processing, data compression, communications.

1.1-D Discrete Wavelet Transforms

The 1-D discrete wavelets transform (DWT), which transforms a discrete time signal to a discrete wavelet representation. The first step is to discretize the wavelet parameters, which reduce the previously continuous basis set of wavelets to a discrete and orthogonal / orthonormal set of basis wavelets.

$$\psi_{m,n}(t) = 2^{m/2}\,\psi(2^m t - n) \quad ; m, n \in Z \text{ such that } -\infty < m, n < \infty \qquad (5)$$

The 1-D DWT is given as the inner product of the signal x (t) being transformed with each of the discrete basis functions.

$$W_{m,n} = \langle x(t), \psi_{m,n}(t) \rangle \quad ; m,n \in Z \qquad (6)$$

The 1-D inverse DWT is given as:

$$x(t) = \sum_m \sum_n W_{m,n} \psi_{m,n}(t) \quad ; m, n \in Z \qquad (7)$$

1.2 2-D wavelet Transform

The 1-D DWT can be extended to 2-D transform using separable wavelet filters. With separable filters, applying a 1-D transform to all the rows of the input and then repeating on all of the columns can compute the 2-D transform. When one-level 2-D DWT is applied to an image, four transform coefficient sets are created. The four sets are LL, HL, LH, and HH, where the first letter corresponds to applying either a low pass or high pass filter to the rows, and the second letter refers to the filter applied to the columns. The Two-Dimensional DWT (2D-DWT) converts images from spatial domain to frequency domain. At each level of the wavelet decomposition, each column of an image is first transformed using a 1-D vertical analysis filter-bank. The same filter-bank is then applied horizontally to each row of the filtered and subsample data. One-level of wavelet decomposition produces four filtered and subsample images, referred to as sub bands. [6] The DCT and DWT shows best result DCT-91% ,DWT-97%.[7]

IV. CLASSIFIER

Artificial neural network and hidden morkov model are the two models which work as classifier.

(1) Artificial neural network:- The basic building block of ANN are processing unit ,topology and learning algorithm[8]. The basic structure of neurons is similar to biological neuron. boltzman principle and Hebbian rule help in the learning of networking of neurons[9].The ANN can

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

be used in pattern recognition. In neural network system every neuron should have to be trained independently. So count of network increases. But the drawback of this system is that as number of input data set varies independently, training of neural network becomes a tedious job [10].

(2) Hidden markov Model(HMM):- The HMM system consists of two phases namely training phase and testing phase. In the training phase, features are extracted using DCT/DWT from the mouth regions and they are given as inputs to estimate the parameters of HMM. Then, the word is recognized in the testing phase. HMMs have been widely and successfully used in speech recognition and handwriting recognition [11] Consequently, they seem to be effective for visual recognition of complex, structured hand gestures such as sign language recognition [12,13].

V. RECOGNITION

Among the existing architecture types, the MLP was selected because of its simplicity and adequacy to solve properly the fault classification problem In this way, the learning database must contain a great variety of faulted scenarios to improve the ANN's Generalization capability. Thus it is essential to use number of simulated fault records to accomplish the ANN's learning. By using this strategy, the ANN can classify correctly simulated and real faults. The most used algorithm to train MLP networks is the back propagation algorithm. However, it converges very slowly and sometimes it may not converge. Levenberg-Marquardt optimization method is used to locate the fault rapidly and correctly. The output of the ANN must indicate which fault type is related to the actual input pattern. Hence, binary coding is used for the ANN's outputs in such a way that a fault is characterized by the presence (1) or absence (0) In order to achieve a good performance for the ANN, many topologies were evaluated. A topology with only one hidden layer with eight neurons showed best fitness for the problem.

VI. RESULT

(Results are obtained making group of even and odd numbers)

Table-1 Digit Recognition Using DWT and ANN Classifier

Test input DWT	Train input DWT (Recognition in %)
Zero	76.66
One	80.00
Two	66.66
Three	56.66
Four	66.66
Five	63.33
Six	73.33
Seven	76.67
Eight	80.00
Nine	93.33

VII. CONCLUSION

Future scope of this project is that we can increase the frequency of the conversion of video to frame and with the help of DWT I would like to state that even though neural networks have a huge potential we will only get the best of them when they are integrated with Artificial Intelligence, Fuzzy Logic and related subjects.

REFERENCES

[1] K.N.Pawar,Y.M.Gaikwad "A visual recognition systems for hearically impaired person –Review"(2014) ,IJETAE

[2] Rachel Ostrand, Sheila E. Blumstein, James L. Morgan 'When Hearing Lips and Seeing Voices Becomes Perceiving Speech: Auditory-Visual Integration in Lexical Access'.

[3] Rowan Seymour, Darryl Stewart, and JiMing 'Comparison of Image Transform-Based Features for Visual Speech Recognition in Clean and Corrupted Videos',Received 28 February 2007; Revised 13 September 2007; Accepted 17 December 2007.

[4] E. D. Petajan, '*Automatic lipreading to enhance speech recognition'*, Ph.D. thesis, University of Illinois, Urbana-Champaign, Ill, USA, 1984.

[5] G. Potamianos and H. P. Graf, "Linear discriminant analysis for speech reading," in *Proceedings of 2nd IEEE Workshop on Multimedia Signal Processing (MMSP '98)*, pp. 221–226, Redondo Beach, Calif, USA, December 1998.

[6] www.mathwork.com

[7] N. Puviarasan , S. Palanivel 'Lip reading of hearing impaired persons using HMM', Department of Computer Science and Engineering, Annamalai University, Annamalainagar 608 002, India.

[8] Ben Krose, Patrick van der Smagt a book on 'An introduction to neural network' Eighth edition , November 1996.

[9] A.K.Jain and Jianchang Mao 'A Artificial neural network –A tutorial'march 1996.

[10] C. M. Bishop a book on 'neural networks for pattern recognition' Clarendon Press OXFORD 1995

[11] A. Kundu, Y. He, P. Bahl, 'Recognition of handwritten word: first and second order hidden Markov model based approach, Pattern Recognition' 22 (3) (1989) 283–297.

[12] J. Yamato, J. Ohya, K. Ishii, 'Recognizing human action in time sequential images using hidden Markov models', Proceedings IEEE Conference on Computer Vision and Pattern Recognition (1992) 379–385.

[13] J. Schlenzig, E. Hunter, R. Jain, 'Recursive identification of gesture inputs using hidden Markov models', Proceedings Second Annual Conference on Applications of Computer Vision (1994) 187–194.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.