

# A static hand gesture classification system for American Sign Language (ASL) fingerspelling and digits

Sunanda Biradar

*Assistant Professor, Department of Computer Science and Engineering,  
BLDEA's VP Dr. P. G. Halakatti College of Engineering & Technology,  
Vijayapur, Karnataka*

Ashwini M. Tuppad

*M. Tech. Student, Department of Computer Science and Engineering  
BLDEA's VP Dr. P. G. Halakatti College of Engineering and Technology,  
Vijayapur, Karnataka*

**Abstract** - Sign language recognition involves hand gesture classification. A method for static hand gesture classification has been proposed for American Sign Language(ASL) fingerspelling alphabets and digits. The system uses skin color based segmentation requiring little post processing. The averages of central moments of order 2 to 9 have been extracted as the features that characterize the hand gestures. The neural network classifier has been employed for recognition and it produced decent classification results of 73.68% with a small feature vector size containing 8 features.

**Key words** – Sign language, static hand gesture recognition, central moments

## I. INTRODUCTION

Sign language recognition has been the theme of machine learning research from many years in the past. The importance of gesture recognition has increased at a very fast pace owing to sign language interpretation attempts as well as the new generation of gesture interface technology[1]. Sign language is mainly a language for the differently-abled deaf-mute community who can't use spoken languages to communicate with others and have difficulty with Web mail[2]. Although they can see, the use of hand signs to interact becomes inconvenient to express one's feelings readily if a common hand sign language is not followed. A standard sign language has a defined set of signs and their meanings, making them easy to understand.

The sign language too uses different gestures for communication mostly the hand. Every sign language is unique with its own vocabulary, hand signs, grammar etc[3]. Sign languages are distinct than spoken languages. The commonality is that they too have many variants like regional languages. It becomes a social problem when interpretation of sign languages becomes tough. The use of hand gestures for interaction either with humans or with machines is comparatively higher than other body gestures like head and eyes due to the fact that hands send more clearer signals and the gestures can be made spontaneously. Gestures are of two kinds- static (also called postures) and dynamic[4]. Static gestures are meant for primary, small signals. Dynamic gestures are used for complex or compound expressions, which are containing smaller parts. Examples are for phrases and sentences. They are done in real-time since these are expressed in a sequence[5,6]. The term 'fingerspelling' refers to the expression of written alphabets of some language through a sign language's gestures.

Static hand recognition is a trending research with wide applications. It has scope in design of human-computer interaction(HCI) interface based on gestures[7,6], robotics[5,7], virtual reality[5], telemedicine and gaming[3]. The aim of our system is however to define a methodology for recognizing static sign language gestures and classify them using less complex recognition technique. We use a small sized feature vector of 8 features for ASL hand gestures.

The paper is organized as follows: Section II presents the literature review on static hand gesture recognition; Section III presents the flow of proposed methodology; Section IV is background on neural network backpropagation algorithm and training algorithms used in the method; Section V contains experimental results followed by conclusion and future scope in Section VI. Finally references are listed.

## II. LITERATURE SURVEY

This section discusses the research carried out in the past in the field of static hand gesture recognition for sign language and few other applications. The survey would highlight the various features considered for proper classification of static hand gestures.

Rokade et al[1] proposed a system for recognizing static American Sign Language(ASL) gestures for numbers from 1 to 9. They have used a novel segmentation technique based on histogram matching in YIQ color space and YCbCr color space mapping[14] skin color pixels to white and background pixels to black. Thinning is employed in order to extract features over thinned image. The recognition of gestures is done by comparing the measures of features extracted namely angle and radial distance. The angle between raised fingers is calculated. The radial distance is the distance from finger tip to end point of wrist. Based on the ranges of these features, raised fingers and hence the gestures of numbers are identified. 92.13% accuracy was seen in recognition.

Priyal and Bora[5] proposed a system for static hand sign recognition that is user and camera location independent. A skin color segmentation method is used and binary silhouettes of gestures considered for feature extraction. The Krawtchouk moments from 0 to 9 orders are extracted and classification is done using nearest neighbour algorithm. Experiments carried out for geometric and Zernike moments also, but the performance was higher in case of Krawtchouk moments.

Zhuang et al[8] proposed a recognition methodology using the concept of compressive sensing. The notion of reconstructing a sample of data from a sparse signal defines it. It uses thresholding of input images and edge detection for obtaining binary hand silhouettes. The combination of two features has been used – local binary pattern(LBP) and curvature scale space (CSS) corner feature. The system is able to cope up with lighting, rotation, scale changes. The overall recognition accuracy was 96.25%.

Gupta et al[6] proposed a gesture recognizer for static hand gestures that makes use of Gabor filter. The method is applied on dataset of binary hand gesture of American Sign Language. The authors have incorporated PCA for feature reduction and LDA for feature extraction. A bank of Gabor filters for 3 scales and 5 different orientation angles is computed. The 15 Gabor filter convolved images are converted to a pattern vector for all 15 filters. Classification is done using a multi – class support vector machine.

Jalilian and Chalechale[2] proposed a novel method for recognizing Persian Sign Language(PSL) hand postures . It uses a segmentation method based on Gaussian model construction for target(skin) pixels and background pixels. Further Bayes's rule is applied for obtaining probabilistic clusters of skin pixels and background pixels thereby extracting the segment of hand part. Radial distance that extends from end point of hand gesture to its centroid together with the Fourier transform are used as the features. Recognition is done using minimum distance classifier where Euclidean distance metric is used. The recognition rate was found to be 95.62%.

Stergioupolou and Papamarkos[4] proposed a novel hand gesture recognizing neural network called Self-Growing and Self-Organized Neural Gas(SGONG) network. They have employed YCbCr color segmentation method stated by Chan and Ngan [13] based on chrominance ranges of skin color distributed in a short interval. Euclidean distances of Cb, Cr values of every pixel in input and the Cb, Cr skin pixel range are computed. If any of the distances is less than the threshold, then pixel is set to white to indicate that it's a hand region pixel. The authors have set the threshold value to 18. Number of raised fingers, palm and hand shape features have been extracted in feature extraction stage. SGONG classifier recognizes based on likelihood based technique and the performance was 90.45%.

Bhavsar et al[9] have proposed a threshold based number recognition system where hand gestures represent numbers from 0 to 10. It makes use of color gloves with different colors on the fingertip locations on the gloves.

The recognition is based on the count of colors on the fingertips, which is compared with the threshold value. Based on certain defined rules, classification of static hand gestures of numbers is completed. It uses the measures of color counts and threshold value. The method restricts the user with two things: the use of color glove and the restriction of background being black.

Reddy et al[10] proposed a gesture recognition system for static as well as dynamic gestures. The authors have used RGB normalization, filtering, adaptive segmentation and other morphological operations in sequence to achieve segmentation. The local orientation histogram is used as the feature for recognizing static hand gestures. The features are found to be robust to lighting changes. The recognition is done using Euclidean distance based minimum distance classification method.

Badi et al[7] have put forth a study of role of machine learning and especially neural network in recognition of hand postures. The authors have taken 6 classes of hand gestures for human computer interaction commands like open, close etc. They have stressed on the accuracy and computational cost of the system used. This decides the usability and applications of the system. Two methods of recognition are implemented- first one using contour of gesture image as features and second using complex moments. Neural network classifier is used for classification by using a back propagation neural network learning algorithm. The accuracy of the methods was 70.83% and 86.38% for first and second method, respectively. They have concluded the scope and suitability of these two methods based on performance, speed and computational complexity.

### III. PROPOSED METHODOLOGY

We propose a methodology to achieve the classification of static ASL hand gestures by extracting the central moment averages of order 2 to 9 and training a neural network classifier to recognize test data. The figure below (Fig. 1) shows the block diagram of the methodology adopted including the different stages involved in the classification process.

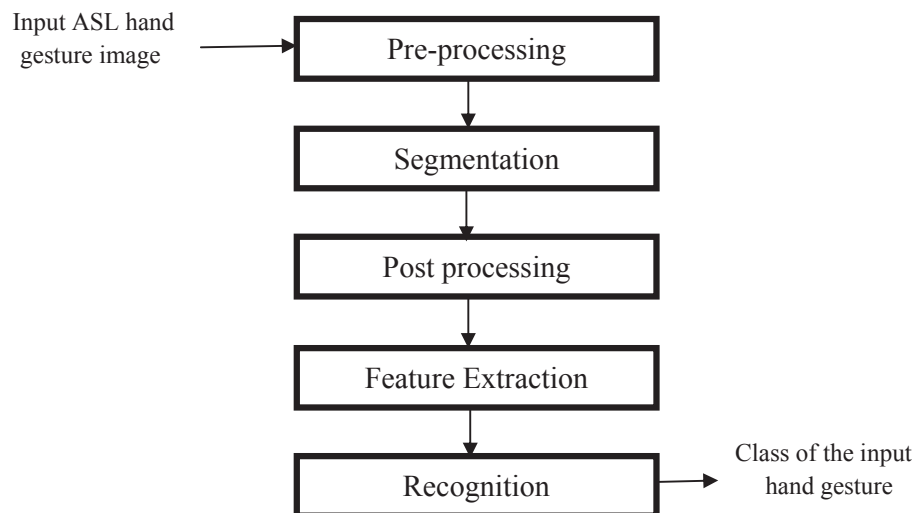


Fig. 1: Block diagram of proposed methodology

The first stage is pre-processing. Here the unprocessed input hand gesture image is subjected to certain steps for quality enhancement. The only pre-processing function carried out here is resizing as the next stage is long and takes time. It has no special input image specifications needed but only image is resized to small size of  $256 \times 192$  pixels to reduce time involved.

The second stage is the segmentation, whose aim is to divide the input image into desirable segments for extracting region of interest or to analyse image information in segments. The method uses the technique proposed by Rokade et al [1]. The steps to be applied to obtain a segmented output can be summarized below:

- I. Convert input RGB image into YIQ form.

- II. Divide the YIQ image into fixed size small windows and compute the histogram of each small window. (The size of each window used was  $4 \times 4$ )
- III. Select a window of the agreed on size which contains skin color pixels and call it as the reference window. Compute its histogram.
- IV. Tabulate the number of times each distinct IQ pair value occurs in each small window and that in the reference window.
- V. Assign the minimum IQ pair value of each window (non-reference) as the minimum among the reference and other non-reference window compared one by one for each IQ pair occurring in reference and non-reference windows.
- VI. Compute the similarity index of each non-reference window as the sum of minimum of all IQ pair repetition values divided by 16. (16 is the total number of pixels in each window and the total number of histogram bins used.)
- VII. The similarity index lies between 0 and 1. Put the similarity values of each window to a new window where each window is represented by a single pixel of value equal to its similarity index. The pixels must be in same row-column format corresponding to the windows. This produces a gray image.
- VIII. Convert input RGB image into YCbCr form. For each pixel in YCbCr image, set the value of pixel to 1 if its Cb component lies the range [80,105] and its Cr component lies in the range [130,165] [1]. Otherwise set it to 0. This is done as it is proved that skin color pixels lie in small color range of Cb – Cr values irrespective of race by Chai and Ngan[14].
- IX. Perform image addition of gray image obtained in VII and output of step VIII. This produces a grayscale output showing skin color pixels in shades of gray.
- X. Binarize the grayscale output of previous step and threshold it taking 3 neighbouring connected windows' similarity indices. This produces the final binary segmented output.

Third stage is post processing in which generally noise is removed and holes are eliminated. Our method employs hole filling operation as part of post processing of segmented image, which contains small holes. Rarely due to shadows there may be extra pixels out of hand gesture region. They can be eliminated by area opening operation, where the group of pixels below a given size are eliminated. The largest connected object in white i.e. hand region is the one whose pixel size is considered for eliminating pixel groups less than its size.

Next stage is the feature extraction where we first compute the central moments of order 2 to 9 for all images considered for training. The output involves the column wise central moment samples for each column in the image. Then we take the average of each column's central moments of the image for order 2 to 9. This produces the required feature vector containing eight floating-point values. Central moments are global shape descriptors. They are the properties of the entire region making up the shape including the boundary rather than considering only the boundary of the shape. The 8 different orders are taken to avoid intersection of feature values for more than one class of inputs and to make distinction.

Last stage is the recognition stage. We have used a neural network classifier to assess the system functionality. A multi-layer feed forward neural network is used and the training feature set contains 190 instances. The training is carried out using different training functions to use a better trained classifier. The testing is carried out on different trained classifiers and the performance of each is noted. The output of testing the trained neural network is the class to which the input ASL hand gesture belongs to. The figure below (Fig. 2) shows the neural network used in the proposed system which uses 8 inputs (I1 to I8) that are 2<sup>nd</sup> to 9<sup>th</sup> order central order moments and the output is the class of input hand gesture. The hidden layer size for which optimal results were observed was 18 among other hidden layer sizes.

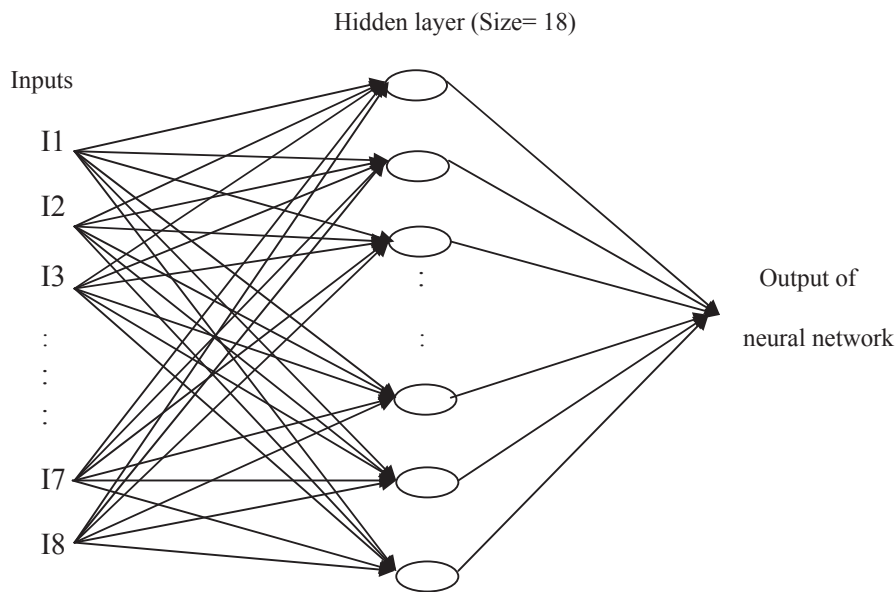


Fig. 2: The structure of neural network classifier used in the methodology proposed that produced recognition rate of 73.68% with hidden layer neurons =18 and training function of Bayesian regularization.

#### IV.

#### NEURAL NETWORK BACKPROPAGATION ALGORITHM

Neural networks are machine learning concepts that imitate human brain functionality[11]. It contains a natural neural network gifted to humans which does all the information processing in parallel and learns it [11]. Artificial neural networks achieve the learning process using a set of parallel, distributed functional units called neurons interconnected using a network of connections. Each connection has an assigned weight. Each network contains at least 2 layers – a hidden layer and an output layer as can be seen in Fig. 2. The inputs to the neural network are a fixed set of parameters that need to be learnt. There exist connections that connect the network inputs to hidden layer neurons but the first layer containing network inputs is not considered a neural network layer. It is because no processing takes place there, only transfer of inputs to neural network occurs. Hidden layer computes intermediate results needed and learns them. The output of a neural network is a layer of output units in case of multiple outputs, which is the result of learning.

Backpropagation is the basic algorithm on which most of the neural network learning algorithms are based. In short, a neural network learns a problem using a set of weights and biases. Considering the problem of hand gesture recognition, the aim of this classifier is to find a set of weights that fit the features of hand gestures[12] extracted and output the class of hand gesture returned by the hidden layer to the output units. The aim of the classifier remains minimizing the error between network outputs and actual outputs called targets. Backpropagation resolves this by using an arbitrary set of weights first [12]. It incrementally improves the result by finding the weights in the steepest direction of network error[12]. This error minimization procedure in the direction of steepest error gradient is called the gradient descent method. The two training algorithms implemented in our system for analysing the results are Scaled conjugate gradient and Bayesian regularization. Conjugate gradient algorithm is a variation of gradient descent algorithm where the minimization of error is done in the direction of conjugates of steepest error gradients. Bayesian regularization is a statistical problem that can address the problem of non-linear function approximation for target functions in pattern recognition problems. Their presumption is that network performance is a function of mean or sum of squares of error between network prediction and actual output.

## V.

## EXPERIMENTAL RESULTS

This section deals with the experimental results starting from segmentation till the final output of recognition. It also presents the results obtained using different training functions and the rate of recognition obtained.

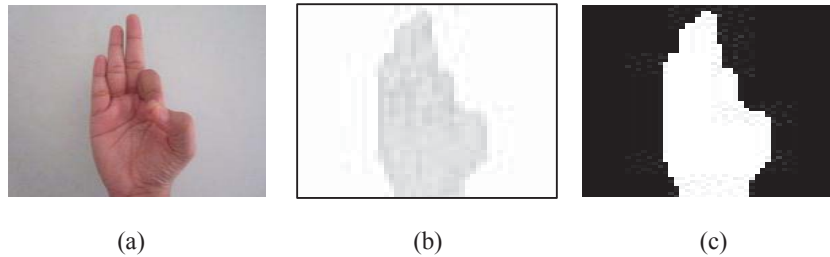


Fig. 3: Output of segmentation – (a) Input RGB image showing ASL sign ‘F’, (b) Grayscale segmentation output and (c) Binary segmented output.

The above figure(Fig. 3) depicts the results of segmentation. Here, there is no need of post-processing as the results are perfect with no noise in them. But there are some cases in which holes were observed. Mostly, many images required little pre-processing. Direct segmentation produced overall good results. Example of one image requiring hole filling after segmentation is depicted in Fig. 4 below.

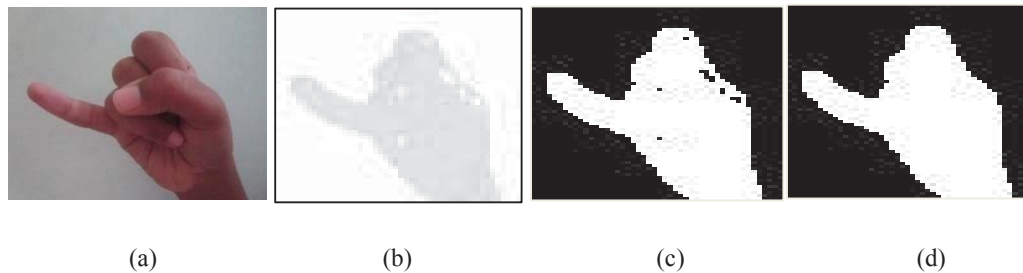


Fig. 4: Segmentation output showing holes in segmented binary hand gesture of letter ‘J’ – (a) Input RGB image, (b) Grayscale segmented output, (c) Binary segmented output containing holes and (d) Final output with holes removed.

The feature extraction results were the feature vectors for each input image in the training set. The test images are classified based on the same features which are input to the classifier during testing and comparison is made between training feature set and test image feature vector. Below is a sample feature vector corresponding to the image(Fig. 5(a)).

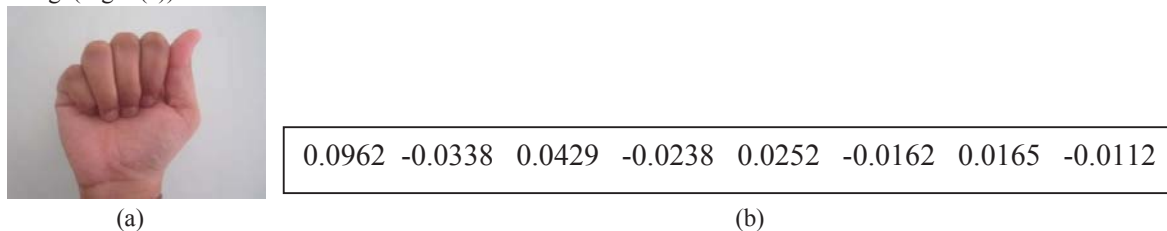


Fig. 5: Example feature extraction result for image of class 1(letter ‘A’) – (a) RGB input image over whose segmented form features are computed. (b) a sample feature vector input image on the left with 8 central moment averages of orders 2 to 9 in sequence.

The recognition of hand gestures is done using neural network classifier and following observations were recorded with different training parameters. The image data collected include 10 alphabets of ASL(A to J) and 9 digits(1-9). The training images included 10 samples for each class.



Training function	Hidden layer size	Epochs	Number of correctly classified gestures (out of 114)	Recognition %
Scaled Conjugate Gradient	16	1000	62	54.38
		800	48	42.10
	17	1000	56	49.12
		800	54	47.36
	18	1000	56	49.12
		800	52	45.61
Bayesian Regularization	16	1000	67	58.77
		800	80	70.17
	17	1000	71	62.28
		800	75	65.78
	18	1000	84	73.68
		800	80	70.17

Table 1: Tabulation of neural network recognition for ASL input hand gestures of alphabets and numbers. The highest recognition rate was 73.68% using Bayesian regularization function, hidden layer neurons=18 and epochs=1000.

The methodology used thus produced a moderate recognition rate of 73.68% against a test dataset size of 114 images of ASL hand signs. The samples for each alphabet/digit for testing collected were 6. The recognition of gestures was better with Bayesian regularization training as the recognition % is recorded in Table 1.

## VI. CONCLUSION AND SCOPE

The proposed hand gesture classification system for ASL hand gestures is able to recognize test data at a moderate recognition rate of 73.68% by extracting averages of central moments from order 2 to order 8. The segmentation method employed takes more computational time as it works block-wise on entire input image. To tackle this, the segmented outputs can be pre-computed and considered as inputs to next stages. This reduces the required time but the technique produces very good segmented outputs with no much pre- and post-processing requirements. The neural network classifier has been used for classification of hand gestures. The results of recognition can be improved in future by extracting more promising features that are well separated as the many classes have to be recognized.

## REFERENCES

- [1] Rokade Rajeshree, Doye Dharmal and Kokare Manesh, "Hand Gesture Recognition by Thinning Method", Proceedings of the IEEE International Conference on Digital Image Processing, pp. 284-287, 2009.
- [2] Jalilian Bahare, Chalechale Abdolah, " Persian Sign Language Recognition Using Radial Distance and Fourier Transform", Image, Graphics and Signal Processing, Vol. 1, pp. 40-46, 2013.
- [3] Wikipedia contributors. (2016, April 26). *Sign language* [Online]. Available: [http://www.en.m.wikipedia/wiki/Sign\\_language](http://www.en.m.wikipedia/wiki/Sign_language)
- [4] Stergiopoulou E. and Papamarkos N., "Hand gesture recognition using a neural network shape fitting technique", Engineering Applications of Artificial Intelligence, Vol. 22, Issue 8, pp.1141-1158, 2009.
- [5] Priyal Padam S., and Bora Prabin Kumar, "A robust static hand gesture recognition system using geometry based normalizations and Krawtchouk moments", Pattern Recognition, Vol. 46, Issue 8, pp. 2202-22019, 2013.
- [6] Gupta Shikha, Jaafar Jafreezal and Wan Ahmad Wan Fatimah, "Static Hand Gesture Recognition Using Local Gabor Filter ", Procedia Engineering, Vol. 41, pp. 827 - 832, 2012.
- [7] Badi Haitham, Hussein Sabah Hasan and Kareem Sameem Abdul, "Feature extraction and ML techniques for static gesture recognition", Proceedings of Neural Computing and Applications, Springer, pp.733-741, 2014.
- [8] Zhuang Huiwei, Yang Mingqiang, Cui Zhenxing and Wang Feng, "Static Hand Gesture Recognition Using Combinational Features and Compressive Sensing", Lecture Notes in Engineering and Computer Science, Vol. 2221 , pp. 376 – 380, 2016.
- [9] Bhavsar Swapna, Futane Pravin and Dharaskar Rajiv V, "Hand Gesture Recognition System for Numbers Using Thresholding", Proceedings of First International Conference on Computational Intelligence and Information Technology, Springer, pp. 782-786, 2011.

- [10] Reddy Sivarajesh K., Swarna Latha P. and Babu Rajasekhara M., "Hand Gesture Recognition Using Skeleton of Hand and Distance Based Metric", Proceedings Of First International Conference Advances in Computing and Information Technology, Springer, pp. 346-356, 2011.
- [11] David Kriesel. (2005, May 27). *A Brief Introduction to neural networks* [Online]. Available: [http://www.dkriesel.com/en/science/neural\\_networks](http://www.dkriesel.com/en/science/neural_networks), pp. 1-12.
- [12] Mitchell Tom M., "Artificial Neural Networks" in *Machine Learning*, INDIAN ed., McGraw-Hill Education, 2013, pp. 81-122.
- [13] Chai D., Ngan K. N., "Face segmentation using skin color maps in videophone applications", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, Issue 4, pp.551-564, 1999.