A Model For Real Time Sign Language Recognition System

P.V.V.Kishore
Dept. of electronics and communications engineering
Miracle Educational Society Group of Institutions
Miracle City, Bhoghapuram,
INDIA

P.Rajesh Kumar
Dept. of electronics and communications engineering
Andhra University College of Engineering
Visakhapatnam,
INDIA

Abstract: **This paper proposes a real time approach to recognize gestures of sign language. The input video to a sign language recognition system is made independent of the environment in which signer is present. Active contours are used to segment and track the non-rigid hands and head of the signer. The energy minimization of active contours is accomplished by using object color, texture, boundary edge map and prior shape information. A feature matrix is designed from segmented and tracked hand and head portions. This feature matrix dimensions are minimized by temporal pooling creating a row vector for each gesture video. Pattern classification of gestures is achieved by implementing fuzzy inference system. The proposed system translates video signs into text and voice commands. The training and testing of Fuzzy Inference system is done using Indian Sign Language. Our data base has 351 gestures with gesture repeated 10 times by 10 different users. We achieved a recognition rate of 96% for gestures in all background environments.**

I. INTRODUCTION

The key principle behind this paper is proposing a method to improve the communication between deaf community and normal people for better social and intellectual life. Sign language is as old as the race itself and its earliest history is equally ambiguous. Sign language is defined as a language of that uses systems of manual, facial and other body movements as the means of communication predominantly among deaf people.

Sign language recognition is multidisciplinary research area involving computer vision, image segmentation, pattern recognition and neural language processing. Sign language recognition is a comprehensive problem because of the complexity of the shapes of hand and head gestures. Sign language recognition requires the knowledge of hands position, shape, motion, orientation and facial expressions. A functional sign language recognition system can be used to generate speech or text making the hearing impaired person more independent. The most complicated part of sign language recognition system is to recognize the simplest hand gestures that must be detected in the image.

M.K. Bhuyan et al. [1] have developed a framework for hand gesture recognition, they propose a gesture recognition system based on object-based video abstraction technique. Their experimental result shows that their recognition system can be used reliably in recognizing some signs of native Indian sign language [13].

Thad Starner et al and others [2, 3, 4] have presented an extensive system which uses a color camera to track hands in real time and interprets American Sign Language (ASL) using Hidden Markov Models (HMMs) with a vocabulary of 40 signs. Signs are modeled with four states-HMM and they have achieved recognition accuracies between 75% and 99%.

Eng-Jon Ong and Bowden [4] have developed an unsupervised approach to train a robust detector the presence of human hands within an image and classified the hand shape. In their work, a database of a hand gesture images were created and clustered into sets of similar looking hands using k-means clustering algorithm. Wilson and Bobick [5] have proposed a state-based technique for the representation and recognition of gesture. Noor Saliza Mohd Salleh et al. [6] have presented techniques and algorithms for hand detection and gesture recognition process using hand shape variations and motion information as the input to the Hidden Markov Model based recognition system. Rini Akmelia et al. [7] have developed a real-time Malaysian sign language translation using colour segmentation and achieved a recognition rate of 90%.

One of the major difficulties arise in the form of background in which signer is located. Most of the methods developed so far use simple backgrounds [8,9,10,11] in controlled set-up such as simple backgrounds, special hardware like data gloves[12,13,14], restricted sets of actions, restricted number of signers, resulting different problems in sign language feature extraction.

The active contours was first introduced by Terzopoulos [15,16]. The basic idea behind active contours is to start with a curve anywhere in the image and move the curve in such way that it sticks to the boundaries of the objects in the image, thus separating the background of the image from its objects. The original snakes algorithm was prone to topological disturbances and is exceedingly susceptible to initial conditions. However with invention of level sets [17] topological changes in the objects are automatically handled.

For tracking non-rigid moving objects the most popular model based technique is active contours [18]. Active contours can bend themselves based on the characteristics of the objects in the image frame. First created active contour models are only capable of tracking an object in a video sequence with a static background [19]. Jehan-Besson et.al [20] tracked objects in successive frames by matching the object intensity histogram using level sets. Though, the tracking error was minimum, it increased if the object experiences intensity transformation due to noise or illumination changes. Almost all active contours methods discussed above suffer from problems related to cluttered backgrounds, lacking in texture information and occlusions from other objects in a video sequence. These problems can cause the performance of tracking non-rigid objects to decrease drastically.

Segmentation and tracking results are fused together to form a feature matrix which is unique from the methods proposed for sign language feature extraction in [21,22,23]. Pattern classification is done by employing fuzzy inference system [24,25,26,27].

The rest of the paper is organized as follows: sect. 2, we introduce the proposed system for recognizing gestures of Indian Sign Language (INSLR). Sect.3, presents proposed segmentation and tracking using active contours. Sect.4, gives the creation of feature matrix. Sect.5, pattern classification using neural network and Sect.6, we discuss experiments under various test conditions. Finally Sect.7, we briefly conclude with future scope of our research.

II.   PROPOSED SYSTEM ARCHITECTURE

The proposed system has four processing stages namely, hand and head segmentation, hand and head tracking, feature extraction and gesture pattern classification. Fig. 1 shows the block process flow diagram of the proposed method. Video segmentation and tracking stage divides the entire frames of images of the signers into hands and head segments and tracking stage calculates the location of the segmented hands and head. Features are extracted and are optimized before saving to the database. This process is repeated for all the available signs using a video of a signer. The sign language database can be upgraded to add new signs into the database.

The fuzzy rule based system is a very powerful tool used by many researchers for pattern classification tasks. Once trained the fuzzy inference system produces a voice and text response corresponding to the sign irrespective of the signer and changing lighting conditions.

III.   SEGMENTATION AND TRACKING

A.   *Hands and Head Segmentation and Tracking*

This section presents the video object segmentation and tracking of hands and head simultaneously from a range of video backgrounds under different lighting conditions with diverse signers. A video sequence is defined as a sequence of image frames $I(x, y, t) : D \rightarrow \Re$ where the images change over time. Alternatively a succession of image frames can be represented as $I_{(n)}$ where $0 \leq n \leq \infty$. The basic principle behind our proposed segmentation and tracking technique is to localize the segment and track one or more moving objects of the nth frame from the cues available from previous segmented and tracked objects in

frames $I_{(1)}, I_{(2)} ..... I_{(N)}$ such that subsequent contours $\Gamma^1, \Gamma^2 .... \Gamma^N$ are available. The sign videos are composed of many moving objects along with the hands and head of the signer. We considered signers hands and hand as image foreground denoted by $I_{(n)}^{f}$ and rest of the objects as the image background $I_{(n)}^{b}$ for the image $I_{(n)}$ in the video sequence. We may denote foreground contour of the hands and head by $\Gamma_{(n)}^{f}$ Our proposed video segmentation algorithm segments hands and head of signers using color, texture, boundary and shape information about the signer given precedent understanding of hand and head shapes from $I_{(n-1)}^{f}$ and $I_{(n-1)}^{b}$.
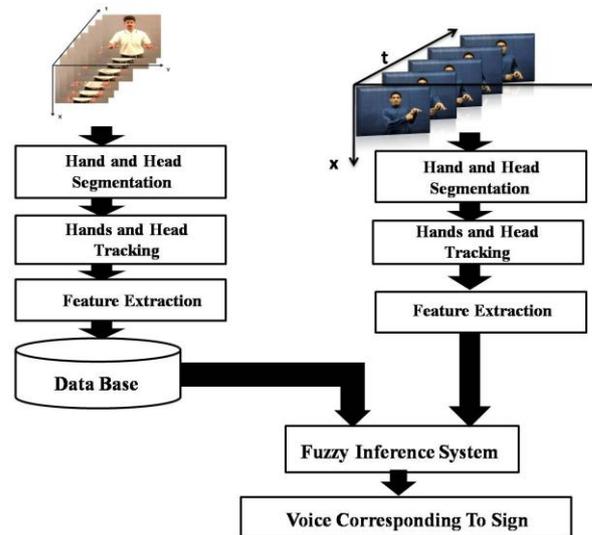

Figure 1.   Sign Language Recognition System Architecture.

B.   *Color and and Texture Formulation*

We choose manually the color plane which highlights the human object from a background of clutter. Once a color plane is identified, texture features are calculated using coorelogram of pixel neighbourhood [28]. Texture is an irregular distribution of pixel intensities in an image. Allam.et.al [29] established that co-occurrence matrix (CM's) produce better texture classification results than other methods. Gray Co-occurrence matrix (GLCM) presented by Haralick.et.al [30] is most effectively used algorithm for texture feature extraction for image segmentation.

Let us consider a color plane of our original RGB video. The R color plane is now considered as a $M \times N$ R coded 2D image. The element of co-occurrence matrix $C^{d,\theta}$ defines the joint probability of a pixel x_i of R color intensity r_i at a distance d and orientation θ to another pixel x_j at R color intensity r_j. For each co-occurrence matrix, we calculate four statistical properties: contrast, correlation, energy and homogeneity.

We used four different orientations $\theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ and two distance measures $d = (-1,1)$ for calculation of GLCM matrix.

Finally a feature vector $f^v(x)$ is produced which is a combination of any one or all of the color planes and texture vector. Thus $f^v(x) = \{f^{(1)}(x), f^{(2)}(x), ..., f^{(n)}(x)\}$ the feature vector contains color and texture values of each pixel in the image. This is a 5D feature vector allocating the first vector for any of the three color planes (R,G,B) and the next four vectors for texture. We can also use all three color planes to represent color, and then the feature vector becomes a seven dimension feature vector.

we classify them as background and foreground pixels using K-Means clustering algorithm. Given dimensional feature vector f^{vect} and the K-Means algorithm classify this vector into categories. The centroids of each group are used to identify each of the clusters where $c = 1, 2, ...N, for \ all \ N > 1$. For every new classification the difference between the new vector and all the centroids is computed. The centroid corresponding to smallest distance is judged as the vector that belongs to the group.

$$d = \min_c \left| f^v(x) - s_c \right| \qquad (1)$$

d is the distance of every new $f^v(x)$ of each frame to the previously computed centroids. All the pixels are classified and average of all pixel values in each cluster is calculated. The centroids are replaced by new average and pixels are classified again until all cluster centers become rigid. In the first frame $I_{(1)}$ all objects and background clusters are created. The object region contains three clusters of foreground $C^{(1)}{}_{f_i}$, where i =1 to 3 and background region $C^{(1)}{}_{b_j}$, where j =1 to 2 into two clusters**.** We will assume at this point is that objects in the video sequence pretty much remain same compared to background that varies due to camera movement or changes in background scenes. This can be taken care by periodically updating the background clusters with some threshold if the changes in consecutive background frame cross the specified threshold.

To move the contour on to the objects of interest we minimize the following energy function E^{CT} from color and texture according to the initial object contour I^f

$$E^{CT}(I^f) = \sum_{i=1}^{3} \int_{obj} D(C^{(n-1)}{}_{f_i}, C^{(n)}{}_{f_i})dx + \sum_{j=1}^{2} \int_{bck} D(C^{(n-1)}{}_{b_j}, C^{(n)}{}_{b_j})dx \quad (2)$$

where $C^{(n-1)}{}_{f_i}$ and $C^{(n-1)}{}_{b_j}$ are object and background centroids of previous frame $C^{(n)}{}_{f_i}$ and $C^{(n)}{}_{b_j}$ are object and background clusters from current frame. The $(n-1)$ frame cluster centroids will become the frame initial centroid and the object contour is moved by minimizing the Euclidean distance between the two centroids. We can implement this by assigning pixel to object in the current frame when $D(C^{(n-1)}{}_{f_i}, C^{(n)}{}_{f_i}) > D(C^{(n-1)}{}_{b_j}, C^{(n)}{}_{b_j})$ and to the background otherwise.

C. Boundary Edge Map Module

we extracted boundary edge map of the image objects which only depends on image derivatives. The way out would be to couple the region information in the form of color and texture features to boundary edge map to create a good tracking of image objects.

We define the boundary B^o as pixels that are present in edge map of the object. The boundary pixels can be calculated by using gradient operator on the image. To align the initial contour Γ^o from previous frame to the objects in the current frame to pixels on the boundary we propose to minimize the following energy function

$$E^O(I^f) = \int_{arc(Length \ of \ obj)} g(B^o(x))dx \qquad (3)$$

where $arc(Length \ of \ obj)$ is the length of the object boundary. The function g is an edge detection function. The boundary energy reaches to a minimum when the initial contour aligns itself with the boundary of the objects in the image. The minimization of energy in eq.3 also results in a smooth curve during evolution of the contour.

D. Prior Shape Information Module

The active contour can be influenced by giving information regarding the shape of the object computed from the previous frames.

The following method in [31] is used to construct the influence of shape of non-rigid objects in the image sequence. As for the first fame $I_{(1)}$ where prior shape information is not available we just use the region and boundary information for segmentation. The segmented objects in frame one are used to initialize contours in the next frame for segmentation and tracking. For $I_{(n)} \forall n > 1$ the tracking of $I_{(n)}$ is given by the level set contour Γ^n which minimizes the energy function

$$E^S(I^f) = \int_{\Gamma^{int}} \phi_0(x)dx$$
$$(4)$$

Thus by applying shape energy to the level set we can effectively segment and track hands and head in sign videos and we could differentiate between object contour modifications due to motion and shape changes.

E. Combining Energy Functions

By integrating the energy functions from color, texture, boundary and shape modules we formulate the following combined energy functional of the active contour as

$$E_{Int}(I^f) = \alpha E_{CT}(I^f) + \beta E^O(I^f) + \gamma E^S(I^f)$$
$$(5)$$

Where α, β, γ are weighting parameters that provide stability to contribution from different energy terms. All terms are positive real numbers. The minimization of the energy function is done with the help of Euler-Lagrange equations and realized using level set functions. The resultant level set formulation is

$$\frac{d\phi^n(x,t)}{dt} = (\alpha R_{CT}(\phi^n) + \beta R^b(\phi^n) + \gamma R^S(\phi^n)) \left\| \nabla \phi^n \right\| \quad (6)$$

where

$$R_{CT}(\phi^n) = -D(p_{in}, \overline{p_{in}}(x)) + D(p_{out}, \overline{p_{out}}(x)) \quad (7)$$

$$R^b(\phi^n) = g(B^O(x)) + \nabla.g(B^O(x)) \left[\frac{\nabla \phi}{|\nabla \phi|} \right] \quad (8)$$

$$R^S(\phi^n) = \phi_0(x) \quad (9)$$

The numerical implementation of eq.6 is computed using narrowband algorithm [32]. The algorithm approximates all the derivatives in eq.6 using finite differences. The level set function is reinitialized when the zero level set clutches the boundary of the object in the image frame.

IV. FEATURE MATRIX

The feature matrix f^{Mat} derived from a video sequence is a fusion of hand and head segments representing shapes in each frame along with their location in the frame. The shape information for an nth frame is present ϕ^n which is a binary matrix equal to frame size. Tracking active contour results in location of hands and head contours in each frame giving their location $(x^{(n)}, y^{(n)})$ values. For an nth frame in the video sequence, the first row of feature vector f^{vect} consists of pixel values with the shape of active contour i.e. the pixels representing segmented head and hand shapes. The second and third row consists of $(x^{(n)}, y^{(n)})$ location information of those pixels that are segmented. The feature vector for each frame is a three dimensional vector f^{vect} representing shape and location information about each segmented hand and head contours. For a entire video in a sign language recognition system, each f^{vect} of three dimension becomes a f^{Mat} feature matrix.

Temporal pooling is engaged to reduce the dimensionality of the feature vector for a particular video sequence. Averaging on each row of the f^{Mat} with respect to number of pixels in a frame will leave us with a new reduced feature vector f^{Nvect} which uniquely represents a particular video sequence. Finally each row in the new feature matrix f^{NMat} consists of feature vector representing each sign video sequence, stored in the database as templates.

V. FUZZY INFERENCE SYSTEM (FIS)

For pattern classification we considered Takagi-sugeno-kang (TSK) or simply sugeno type Fuuzy inference system because the output membership functions are linear or constsnt. Sugeno fuzzy inference system consists of five steps, fuzzification of input variables, applying fuzzy 'and' or 'or' operator, calculating the rule weights, calculating the output level and finally defuzzification. Many methods are proposed to generate fuzzy rule base. The basic idea is to study and generate the optimum rules needed to control the input without compromising the quality of control. In this paper the generation of fuzzy rule base is done by subtractive clustering technique in sugeno fuzzy method for classification video. Cluster center found in the training data are points in the feature space whose neighborhood maps into the given class. Each cluster center is translated into a fuzzy rule for identifying the class. A generalized type-I TSK model can be described by fuzzy IF-THEN rules which represent input output relations of a system. For multi input single output first order type-I TSK can be expressed as

IF xi is Q_{lk} and x_2 is Q_{2k} and ... and x_n is Q_{nk}, THEN Z is

$$w = p_0 + p_1\chi_1 + p_2\chi_2 + ... + p_n\chi_n$$

(10)

Where x_1 , x_2.... x_n and Z are linguistic variables; Q_{1k}, Q_{2k} ...and Q_{nk} are the fuzzy sets on universe of discourses U., V . and W, and p_0 ,p_1,...p_n are regression parameters.

With subtractive clustering method, X_j is the jth input feature x_j ($j \in [1, n]$), and Q_{jk} is the MF in the kth rule associated with j^{th} input feature. The MF Q_{jk} can be obtained as

$$Q_{jk} = e^{\left[-\frac{1}{2} \left(\frac{X_j - X_{jk}}{\sigma} \right)^2 \right]}$$

(11)

Where x_{jk} is the jth input feature of x_k, the standard deviation σ of Gaussian MF given as

$$\sigma = \sqrt{\frac{1}{2a}}$$

(12)

VI. RESULTS

In the first experiment we started with a video sequence which is shot in a lab using web cam with dark background and with an additional constraint that signer should also wear a dark shirt. This video sequence is part of the database we have created for sign language recognition project. Our Sign language database consists of 351 signs with 8 different signers. The frame size of the 320×480 Fig. 2 shows where we run our segmentation and tracking algorithm with values of $\zeta = 0.22, \eta = 0.47$ and $\chi = 0.2$ The object and background clusters are made of three and two clusters.

The experiments are performed in R color plane. As such we can do it any color plane or with color videos. The problem with full color sequences pertaining to sign language is that sign language videos contain large sequence of frames with lot of information to be extracted. Fig. 2(a) shows four frames from a video sequence of the signer performing a sign related to English alphabet 'X'. This simple sign contain 181 frames. Fig. 2(b) shows the results obtained from our proposed method. The inclusion of prior shape term in the level sets segments that the right finger shape in spite of being blocked by the finger from left hand. This segmentation and tracking result will help in good sign recognition. Fig. 2(b) and 2(c) shows the effectiveness of the proposed algorithm against the Chan-Vese (CV) model in [33].

Figure 2. Experiment one showing our proposed segmentation algorithm on sign language videos under laboratory conditions. Frames 10, 24, 59,80 are shown. Row (a) shows four original frames. Row (b) shows the results from proposed algorithm and Row (c) Shows results of CV algorithm in [33].

We also experimented with more real time scenarios such usage of sign language in offices or colleges, so that sign language recognition system can be implemented under real time. The video sequence that is considered is taken in a room where due to less cabin space the signer hands cover the face of the signer in most of the occasions. Hence it is difficult to segment and track the hands or the face of the signer. For this image sequence we have manually extracted the signer's hands and head portions from the first frame $I_{(1)}$ which is used to initialize the proposed level set. The weighing parameters in eq.27 $\alpha = 0.12$, $\beta = 0.25$, $\gamma = 0.66$, we observed that segmentation is good if shape term in eq.6 weight is increased. Because in this real time video the color and texture information does not reveal much of information. Similarly the boundary information also provides insufficient data under the influence of such a background clutter. Fig. 3 Shows the results of tracking in the above discussed scenario. Tracking error plot shows our method tracks good in real time environments compared to method in [33].

Figure 3. Experiment Shows segmentation and tracking in closed room with cluttered background. (a) Original video sequence of frames 120, 122,125,128, (b) Segmentation and Tracking using method in [33] (c) Segmentation and Tracking using proposed method.

Classification of different gestures is done using Fuzzy inference system for 351 different gestures of Indian sign language by 9 different signers. Table I gives details of the fuzzy inference system used for gesture classification. The performance of the system is evaluated based on its ability to correctly classify signs to their corresponding speech class. The recognition rate is defined as the ratio of the number of correctly classified signs to the total number of signs:

$$\text{Re} cognition\ Rate(\%) = \frac{Perfectly\ Classified\ Signs}{Total\ Signs\ Used} \times 100$$

The total number of signs used for testing is 351 from 10 different signers and the system recognition rate is close to 96%. The system was implemented with MATLAB version 7.0.1.

Table I.　　Details of fuzzy inference system used for gesture classification.

Fuzzy Model Parameters	Method
Name	'fis_inslr'
Type	sugeno'
And Method	'min'
Or Method	'max'
Defuzz method	'wtaver'
Imp Method	'prod'
Agg Method	'sum'
Input	[1x80 struct]
Output	[1x1 struct]
Rule	[1x25 struct]

Classification of different gestures is done using Fuzzy inference system A graphical User Interface (GUI) has been created to automatically train and recognize the gestures as shown in the fig. 4 and fig. 5.The GUI has two parts, a training part and a

Figure 4.　Demo of Gesture sign 'FIVE' with signer under complex background.

VII.　CONCLUSIONS AND FUTURE WORK

Our proposed techniques try to make the sign language recognition system a close reality. The method proposed combines effectively the color, texture, boundary and prior shape information to produce an effective video segmentation and tracking of sign language videos under various harsh environments such as cluttered backgrounds, poor lighting, fast moving objects and occlusions. The color and texture information is extracted statistically by creating a feature vector and classifying each pixel in the frame to object and background pixel. Boundary information is provided by the divergence operator along with the curvature of the object under consideration. Including Shape information from the previous frame did a whole lot of difference to the level set minimization to segment correctly and track effectively the occulted hand from other hand and also head some times. We have effectively demonstrated by experimentation of the proposed method by applying it to video sequences under various conditions. Nevertheless there are challenges which are to be addressed to apply this method to continuous sign language recognition systems to carry out the recognition in real time. The

recognition part. Training is used to train new signs and store them in the data base of sign language system. Recognition tries to recognize the trained videos along with unknown videos of different signers. Training consists of four processes that can be activated by the use of buttons Load Video, Extract Features, Save Features along with the text in the text box and finally train button to finish the training. View FIS shows the user fuzzy inference system used. Once the load video button is pressed the video is played on top portion of the training side which is interfaced through Microsoft media player.
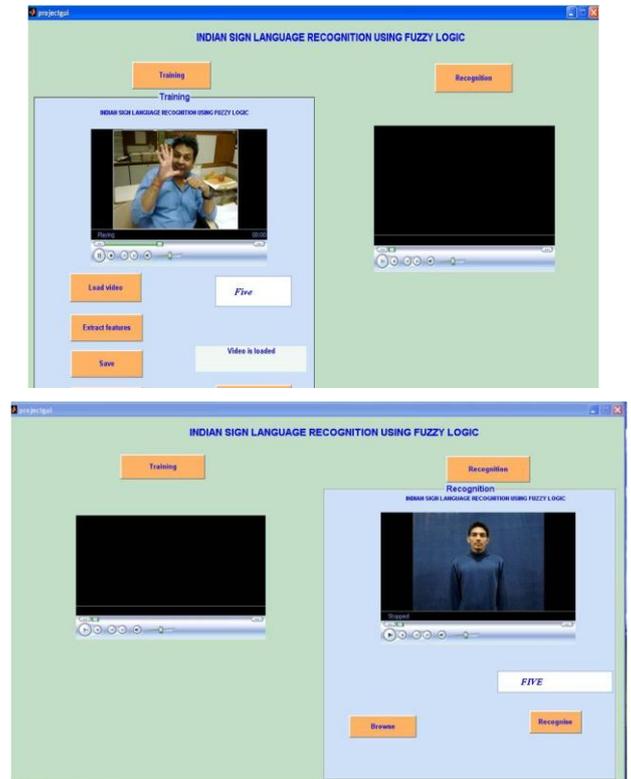

Figure 5.　Testing Demo of Gesture sign 'FIVE' with a different signer with simple background.

recognition rate of the overall sign language recognition system is 96% which is a good score for our proposed real time system.

VIII.　REFERENCES

[1] M.K Bhuyan, D. Ghoah and P.K. Bora (2006), 'A Framework for Hand Gesture Recognition with

[2] I. M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Syst., Man, Cyben., vol. 22, no. 6, pp. 1414-1427, 1992.

[3] Allili, M.S., Ziou, D.,(2006), *"Automatic color-texture image segmentation by using active contours"*. In: Proceedings of 1st IEEE International Workshop on Intelligent Computing in Pattern Analysis/Synthesis, Xi'an, China, 26–27, LNCS 4153, pp. 495–504.

[4] S. Allam, M. Adel, P. Refregier, "Fast algorithm for texture discrimination by use of a separable orthonormal decomposition of the co-occurrence matrix," Applied Optics, vol.36, 1997, pp.8313–8321.

[5] R. M. Haralick, K. Shangmugam, I. Dinstein, (1973) *"Textural Feature for Image Classification,"* IEEE Trans on Systems, Man, Cybernetics, 3(6), pp.610—621.

[6] Allili, M.S., Ziou, D.: Automatic color-texture image segmentation by using active contours. In: Proceedings of 1st IEEE International Workshop on Intelligent Computing in Pattern Analysis/Synthesis, Xi'an, China, 26–27, August 2006, LNCS 4153, pp. 495–504.

[7] Ketut Fundana · Niels C. Overgaard · Anders Heyden,(2008), "Variational Segmentation of Image Sequences Using Region-Based Active Contours and Deformable Shape Priors" Int J Comput Vis vol.80: pp289–299.

[8] T. Chan and L. Vese(2001) *"Active contours without edges"*. IEEE Transations on Image Processing,Vol. 10 No.2,pp266-277.

[9] J. Mackie ,B. McCane, "Finger Detection with Decision Trees", University of Otago, Department of Computer Science.

[10] Claudia N¨olker, Helge Ritter, "Detection of Fingertips in Human Hand Movement Sequences", Gesture and Sign Language in Human-Computer Interaction, Proceedings of the International Gesture Workshop, Published by Springer, 1997, pp. 209-218.

[11] E. Stergiopoulou, N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique", Engineering Applications of Artificial Intelligence (ELSEVIER), vol. 22,pp.1141–1158, 2009, doi: 10.1016/j.engappai.2009.03.008.

[12] Mehdi, S.A.; Khan, Y.N. "Sign language recognition using sensor gloves" Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02. Volume 5, pp18-22, Nov. 2002.

[13] W.Kadous. Recognition of Australian Sign language using instrumented glove. Master Thesis. University of New South Wales. October 1995.

[14] Tan Tian Swee, A.K. Ariff, Sh-Hussain. Salleh, Siew Kean Seng, and Leong Seng Huat, *"Wireless Data Gloves Malay Sign Language Recognition System"* Proceedings of the international conference on Information and computing Systems, 2007, ICICS'07,pp56-59.

[15] D. Terzopoulos and K. Fleischer (1988) *"Deformable models"*. The Visual Computer, Vol.4 No.6, pp306-331.

[16] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer.(1987) *"Elastically deformable models"* In Comp.Graphics Proc., pages 205-214. ACM Press/ ACM SIGGRAPH.

[17] Osher, S., Sethian, J. (1988) Fronts propagating with curvature-dependant speed: algorithms based on Hammilton-Jacobi formulations.J. Comput. Phys. Vol.79 No.1, pp12–49.

[18] L. Brkthes, P. Menezes, E Lerasle and J. Hayet "Face tracking and hand gesture recognition for human-robot interaction" Pmcwdingr of the 2004 IEEE InternatIona1 Confennee on RobotIco 6 Automation New Orleans. April 2004.

[19] Mansouri.A.R, Region tracking via level set pdes without motion computation. IEEE Trans. Pattern Anal. Mach. Intell. 2002, Vol.24, no.7, 947–961.

[20] Jehan-Besson, S, Barlaud, M.,Aubert,G.: Detection and tracking of moving objects using a new level set based method. Proceedings of IEEE International Conference on Pattern Recognition, Barcelona, Spain, 3–8 September, 2000, pp. 7112–7117.

[21] Newby, G. (1993). Gesture recognition using statistical similarity. In Proceedings of virtual reality and persons with disabilities.

[22] Birk, H., Moeslund, T. B., & Madsen, C. B. (1997). Real-time recognition of hand alphabet gestures using principal component analysis. In Proceedings of the 10th Scandinavian conference on image analysis.

[23] Rangarajan, K., & Shah, M. (1991). Establishing motion correspondence. CVGIP: Image Understanding, 54, 56–73.

[24] M. Sugeno, "An introductory survey of fuzzy control," Inform. Sci., vol. 36, pp. 59-83, 1985.

[25] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller-Part-I and Part-II,"IEEE Trans. Syst., Man, Cybem., vol. 20, no. 2, pp. 404-435, 1990.

[26] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Trans. Syst., Man, cybern., vol. 15, no. 1, pp. 116-132, 1985.

[27] L. X. Wang and I. M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Syst., Man, Cyben., vol. 22, no. 6, pp. 1414-1427, 1992.

[28] Allili, M.S., Ziou, D.,(2006), *"Automatic color-texture image segmentation by using active contours"*. In: Proceedings of 1st IEEE International Workshop on Intelligent Computing in Pattern Analysis/Synthesis, Xi'an, China, 26–27, LNCS 4153, pp. 495–504.

[29] S. Allam, M. Adel, P. Refregier, "Fast algorithm for texture discrimination by use of a separable orthonormal decomposition of the co-occurrence matrix," Applied Optics, vol.36, 1997, pp.8313–8321.

[30] R. M. Haralick, K. Shangmugam, I. Dinstein, (1973) *"Textural Feature for Image Classification,"* IEEE Trans on Systems, Man, Cybernetics, 3(6), pp.610—621.

[31] Allili, M.S., Ziou, D.: Automatic color-texture image segmentation by using active contours. In: Proceedings of 1st IEEE International Workshop on Intelligent Computing in Pattern Analysis/Synthesis, Xi'an, China, 26–27, August 2006, LNCS 4153, pp. 495–504.

[32] Ketut Fundana · Niels C. Overgaard · Anders Heyden,(2008), "Variational Segmentation of Image Sequences Using Region-Based Active Contours and Deformable Shape Priors" Int J Comput Vis vol.80: pp289–299.

[33] T. Chan and L. Vese(2001) *"Active contours without edges"*. IEEE Transations on Image Processing,Vol. 10 No.2,pp266-277.