Hand Gesture Recognition using Depth Estimation for Indian Sign Language

Shailendra Badwaik, Shashikant Lokhande

Dept. of Electronics and Telecommunication, Sinhgad College of Engineering, SPPU, Pune, India

ABSTRACT-Hand gesture recognition sign languages like American, British, Arabic, etc. are explored by researcher to great extent. Indian Sign Language (ISL) is most appealing area for research in India and nearby countries. On a contrary ISL is standardized in March 2017 by government of India and the research work for recognizing ISL hand gestures is in wee phase. Proposed Hand Gesture Recognition System (HGR) uses vision based approach of Human Computer Interaction (HCI). Interactive Graphical User Interface (GUI) is developed for HGR system to train new hand gestures.This system is signer independent. In training stage, hand gestures are trained under specific label by evaluating parameters like solidity, circularity, Minimum Bounding Rectangle (MBR), etc.Theseparametersare evaluated using binary hand gesture image. YCgCrcolour space model is used to segment hand region from input image, it is robust against scaling and rotation. In testing stage, parameters evaluated in real time and classified by Naïve Bayes classifier to specific label. Lucas-Kanade optical flow method is used for dynamic gesture recognition. Background subtraction algorithm implemented to get rid of complex and skin color saturated backgrounds, it is applied in both training and testing stage. Inculcating background subtraction and different set of feature improved system performance. Provision of the depth warning message, illumination condition makes system more robust. Accuracy for numbers gesture is 96% and for five static ISL gesture accuracy is near about 92%. For dynamic gesture, singer should be careful while doing gestures because small wrong motion lead to wrong interpretation.

Keywords-*Background subtraction; convex hull; Indian Sign Language (ISL); Naïve Bayes classifier.*

INTRODUCTION

Sign language predominantly used by thehard in listening or people who can hear but unable to speak. According to census in 2016, 20 million people in India are speech impaired of 1.2 billion[1]. By the census of India in 2016, speech impaired population is 9% in total disabled population in India is shown in figure 1 by pie chart.Sign language communication involves two types of signals for sign. These are manual and non-manual signals to do gesture sign. Manual signscomprisesarms, hands and fingers, whereas non-manual signs consistbody, head, face and eyes. These gestures are practice by the hard in listening people for communication but normal people are reluctant to learn sign language. This creates communication gap between the hard in listening peoplewith the normal people.

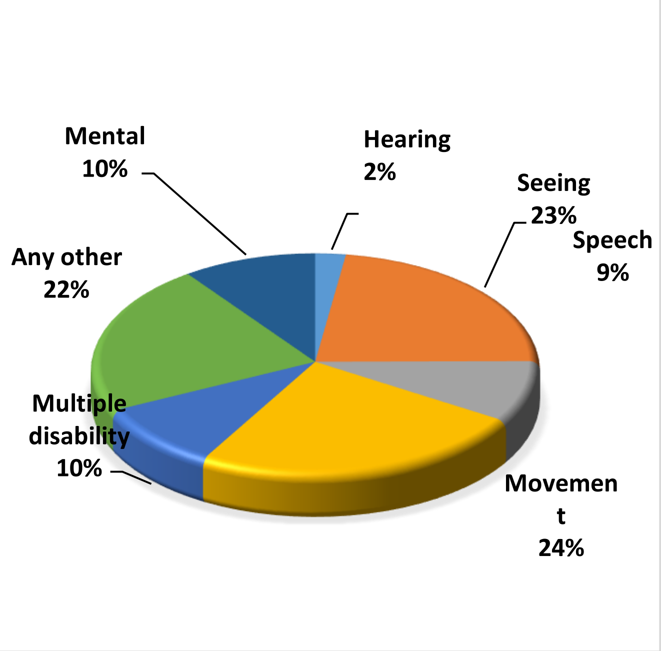


Fig. 1 : Disabled population in India

Sign language interpreters may help deaf people but this is very costly and not feasible solution. Automatic sign language gestures recognization system is needed to overcome commination gap between hard in listening people and normal human being. Indian Sign Language (ISL) is used by in India, Sri Lanka, Nepal, Bangladesh, and border regions of Pakistan.

Glove/Device based and vision based are two main techniques used in the sign language recognition. Glove/Device based gesture recognizationapproach are costly and not feasible for user as user'sgesturemoments are restricted.Use of vision based approach reduce the processing time and boost the recognition accuracy. To detect and track human hands skin colour plays a prominent role. This work gives real-time performance as it encompasses skin area segmentation using the YCgCr color model. Background subtraction algorithm reduces the complex problem of the skin like object and also enhances accuracy of recognition. Implemented algorithm, firstly find the counter for the gesture captured and crop the hand region by making rectangle around it. Biggest binary linked object (BLOB) algorithm will remove the small and non-linked patches. Getting biggest binary linked object convexity defect algorithm employed (followed by convex hull) to compute finger count, convex area, solidity, circularity, etc. These evaluated parameter unique for every gesture. Naïve Bayes classifier trained with these parameters in training stage. In testing stage, hand gesture of deaf and dumb human captured in real time. The statistical parameter evaluated as like in training stage and fed to Naïve Bayes classifier which classify it to particular class. Gesture audio interpretation is produced inHindi / Marathi / English for normal human.

LITERATURE SURVEY

There are mainly two categories for Human Computer Interaction (HCI). First category lies under wearable electronics as it uses data glove for hand gesture recognition and second category isvision based hand gesture recognition[2],[3]. Data glove based system collect digitized data of hand, finger motion by sensor mounted on the fingers and hand. Though such devices provide fast and very accurate motions of the finger and hand, but sensors used are very expensive. Another problem with data gloves is wires connected to each sensor makes user restless. Whereas, vision based approach not demand physical contact with the computer so it provide more viability to user. Vision based hand gesture recognition is broadly classified into two categories. First is, 3D model based and second is, appearance based. The 3D hand model,works on the 3D kinematic hand model. The comparison between the input images and the 2D appearance reflected due to projected 3D hand model, predicts the hand parameters. Though this model can identify hand gestures more effectively but it is computationally expensive. In real time scenario, appearance based HGR models works well. Appearance based models extract features from 2D images which infer the gesture directly compared with frames from the live video.

System for New Zealand Sign Language is developed in[4], which tracked 13 gestures with bare hand. Developed hand gesture tracking clearly recognizes static posture of hand and fingers. More accurate result for posture obtained by using markers with high-resolution input images. The problem with this system is high computational cost, as they uses high-resolution images.In[5], Dardas and Georganas developed real time system to interact with application through hand gesture. They developed an algorithm to segment bare hand in cluttered backgroundby skin detection and posture contour. Scale invariance feature transform (SIFT) extracts keypoints in every training image to train multiclass Support vector machine. Extracted keypoints are quantized using k-means clustering and further they are map into unified dimensional bag-of-words vector. These bag-of-words are input for building the multiclass SVM. Though their system have satisfactory real time performance, but it only track and recognize static postures.In[6] Muhammad RizwanAbid*et al* developed system for smart home using dynamic sign language recognition. The Developed system consist two parts: First, image processing (IP) and second, stochastic linear formal grammar (SLFG). IP module recognizes individual gesture using bag-of-features and local part model approach. 3-D multiscale whole-part features extracted by dense sampling and represented by 3-D histograms of a gradient orientation descriptor. The bag-of-words achieved by k-meansclustering are input for building the nonlinear SVM. Syntactically validity of sentences of sign language is analyzed by SLFG. Though developed system works on sentences of dynamic signs, its real time performance is not satisfactory because of SLFG module. In [7]QiongFei*et al* developed HGR system on Q6455 DSP board. They employed two different approaches for static gestures (hand posture) and dynamic gestures. An algorithm to computes seven different moments and based on that hand posture recognition is done. Also optical flow is applied for tracking and direction encoding of dynamic gesture. Robustness achieved by implementing highly reliable algorithm on DSP board. Real time performance achieved by using four dedicated versatile signal processing module TI-TMS320C6455. On the basis of our experimental observations and analysis we find out moments are not candidates in real time HGR. In [8]S. C. Badwaik*et al*developed system for the deaf and dumb people which recognizes and captured ISL gesture and gives audio output. In this previous work we tracked the gesture using Kalman filter which works in two steps: First step is to predict and second step is to update. Features for tracked gesture are extracted using SIFT. In training stage features for the individual ISL word is stored. In testing Extracted features are matched with stored features, the gesture with correct match is recognized and audio output is generated. Real time hand gesture direction tracking is not satisfactory in this work.

SYSTEM OVERVIEW

The ISL recognition systemisdevelopedasshown in fig. 2, itcomprisesof two phases: First phase isofflinetraining stage and second is real time training stage.

A. Training stage :

Hand gesture images are captured by camera, background subtraction algorithm employed extract the foreground object and it is pre-processed. Pre-processing involves RGB to YCgCr[9] to binary conversion and further morphological operations, BLOB analysis, and optical flow for the dynamic gestures. The developed system works on the contour based algorithms so contour parameters like MBR, circularity, rectangularity, counter axis angle, convex area are extracted.Based on this training stage is build using Naïve Bayes classifier. The testing is done by capturing ISL hand gestures in real time by web camera and classified. The training stage is offline stage in which classifier is built for real time classification. At the start of training stage background image is captured and further model for the background subtraction use this image. Logitech C270 HD webcam is used to capture the background image and hand gestures. Training stage is offline stage and do not put any constraint.

1. *Background subtraction algorithm*

Mixture of Gaussian (MOG) is mostly used background subtraction algorithm [9]. MOG can be approached in two ways

1) Temporal pixel-wise MOG model
2) Spatial MOG model

In this paper, first approach of MOG i.e. temporal pixel-wise MOG model is used to model the background and extract the foreground hand object.

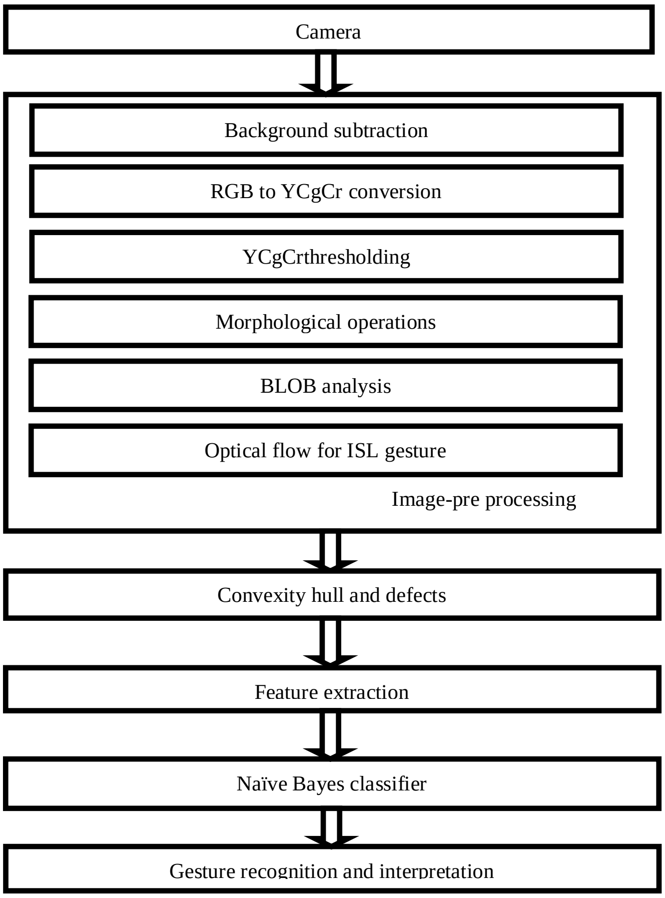


Fig 2: System Block diagram

2. *Skin segmentation*

The process of extracting the skin colored objects from the captured image is skin segmentation.In sign language recognition hand segmentation is key task.Skincolour segmentation is used to extract hand gesture from background. Region of interest hand is segmented from Skin color is a premier parameter for hand localization and segmentation. Colour based segmentation are invariant to geometric transformations and are computationally simple. YCgCr is the best color space model to segment the skin colored objects. Image captured by webcam is in RGB format. So captured image is firstly converted into YCgCr image and then threshold segmenting method applied to get skin segmented out. Threshold for the skin color in YCgCr color space model are given in equation(1).

$$(Y > 40)\ AND\ (105 \geq Cg \leq 138)\ AND\ (145 \geq Cr \leq 195) \quad (1)$$

Pixel falling in the above range is indicated by binary ones and remaining are binary zeroes. With specific threshold for Binary linear Object (BLOB) area only hand gesture BLOB is separated.Furthermorphological operations and BLOB analysis is carried out. Morphology operations consist of two basic operations, dilation and erosion, these operations are backbone of the morphological operations and any other named operations are derived from those. Binary foreground object image subjected to morphological operations to achieve clear noise free, linked contour. By applying thresholding on the BLOBs hand region is extracted.

3. *Optical flow*

Optical flow computes the motion between or sequence of frameswithoutknowing the content of those frames. When object lie in the previous frame and the current frame. In this context assigning velocity to each pixel,called as dense optical flow. Such approaches leads to the high computational cost which put burden on the real time performance of the system. So the alternative option is sparse optical flow. In this algorithm region of interest (ROI) is observed for flow. These points have quantified with certain desirable properties that makes tracking relatively robust and reliable. In proposedsystem Lucas-Kanade (LK)[11] tracking technique is used. For faster motion Pyramidal approach of Lucas Kanade is used. The basic idea of the LK rely on three assumptions.

i)*Brightness constancy:* the brightness of a pixel is unchangedin successive frames

ii)*Temporal persistence/small movements:* With respect to time image motion surface patch

changes slowly.

iii)*Spatial coherence:* in a scene,nearby points belong to the same surface, havesame

motion, and project to adjacent points on the image plane.

4. *Convexity hull and convexity defects*

The convex hull[12] in Euclidean space is set of small points consist all set of given points. Convex hull can be any polygon containing all set of points. It is straight lines drawn around contour region covering all the hand area. Convex hull for the hand is shown in figure 3. By convex hull is depicted close to the contour of the hand, which accommodates set of contour points of the hand among the hull. Convex hull employs minimum points to form the hull to reconcile all contour points inside or on the hull to conserve the property of convexity. So the defects in the convex hull represented with respect to the contour lined on hand. In case the contour of the object is outside from the convex hull, defect are found. Convexity defects are represented in vector form, comprises begin and ending points of the line of defect in the convex hull. Vector incorporatestwo main facts: first is index of the depth point in the contour and second is its depth value from the line.

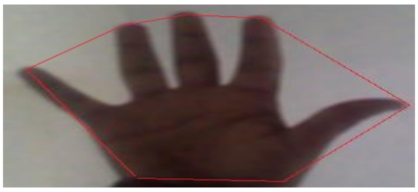


Fig. 3: Convex hull of hand posture

5. Feature Extraction

To train the classifier every gesture should be quantified with some unique features. In this paper we used four parameter to define a gesture viz. Circularity ratio, solidity, convex area, rectangularity, and finger count for static gesture. For dynamic gesture hand direction estimated by LK optical flow is considers along with five parameters in static gestures. Features are evaluated on binary image hand contour as follows:

a)Circularity ratio[13]:

CR defines how Shape is closer to the circle with some arbitrary radius depend on the contour shape, given equation (2).

$$\textit{Circularity ratio} \;=\; A_s \,/\, A_c \quad (2)$$

Where, A_s is area of shape and A_c is area of the circle (same perimeter)

b) Solidity[13]:

How much shape is convex or concave is defined by solidity given in equation (3)

$$\textit{Solidity} \;=\; A_s \,/\, H \; (3)$$

Where, H convex hull area of shape

c)Rectangularity:

Rectangularity defines how much shape is closer to rectangle, formulated as in equation (4)

$$\textit{Rectangularity} \;=\; A_S \,/\, A_R \; (4)$$

Where, A_R is area of MBR

d) Minimum Bounding Rectangle (MBR):

The MBR or envelope, is a bounding rectangle with $2-D\ (x,\ y)$ coordinate system, as $min(x),\ max(x),\ min(y),\ max(y).$

$$\textit{MBR area} = \left(max(x)-min(x)\right)\left(max(y)-min(y)\right)$$

(5)

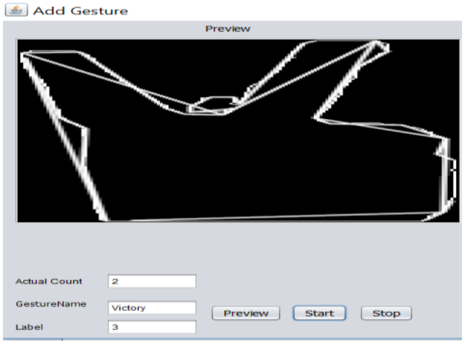
e)Contour orientation / Angle

Hand Contour orientation is obtained by combining second order moments of the contour along the x -axis.

$$\textit{Angle} \;=\; 2.\, m_{11} \,/\left(m_{02}-m_{20}\right)(6)$$

f) Finger Count

By using convexity hull and convexity defects, tip and valley points are computed. By using these points finger count is evaluated



Add Gesture

Preview

Actual Count 2

GestureName Victory

Label 3

Fig. 4 : "Actual count" tab

To avoid wrong training "Actual count" tab is provided. User should manually enter the value of the finger count in "Actual count" tab. Correct finger count can be observed and same can be entered by user while training hand gestures data if it is wrongly interpret figure count . This will avoid deleting such training set with wrong finger count by providing fingers "Actual count". Actual count tab is designed and its GUI isshown in figure 4.

6. Naïve Bayes classifier

The goodness of NaïveBayes classifier[14] is strong Naïve (independence) between the features. This classifier is probabilistic classifierdesigned by integrating Bayes' theorem. In proposed work, features like solidity, circularity ratio rectangularity, etc. are strongly independent of each other. Also single Naïve Bayes classifier can easily classify between multiple classes.

Probabilistic model

Consider gesture defined with n features

$$X \;=\; \left(x_1,\ x_2,\ x_3 \ldots x_n\right),$$

n = no. of features and

k = no. of classes

C_k = class label

By incorporating Bayes' theorem[15], the conditional probability isgiven as

$$P\left(C_k \mid X\right)=\frac{P\left(C_k\right)P\left(X \mid C_k\right)}{P\left(X\right)}$$

Above equation can be written in Bayesian probability terminology as

$$Postrior = \frac{Prior \times Liklihood}{Evidence} \qquad (8)$$

Naïve Bayes classifier gives possibility of gesture X under label C_k. Larger posterior probability for label C_k , more chances of X is fall under label C_k.

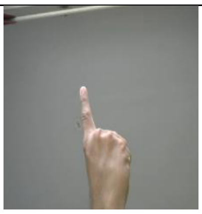
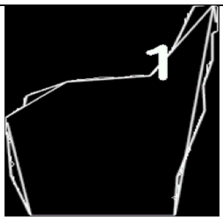
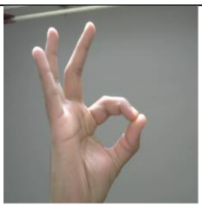
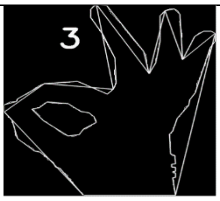
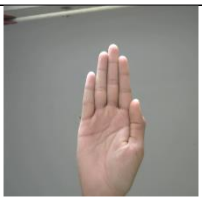
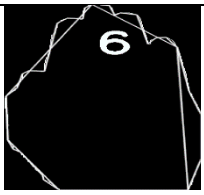
B. Testing stage :

Training stage is offline stage so processing time is not major constraint while doing training. In testing stage to achieve real time performance is verified. Frames captured from webcam after the background subtraction are passed to the hand segmentation. Feature extracted are feed to the Naïve Bayes classifier to recognize hand gestures. Gesture is classified in specific label by matching features and audio output is played corresponding to the label recognized.

EXPERIMENTAL RESULTS

We used standard Indian sign language database for testing of developed system. Table I shows experimental results of static ISL gestures. Static gestures candle, good, stop and dynamic gesture minus is shown in this table.

Table I. Static and Dynamic ISL Gesture Recognition

Gesture Name	Original Image	Recognized Gesture
Candle		
Good		
Stop		

Our main contribution is addition of the background subtraction algorithm andimplementation of the LK optical flow algorithm that work for dynamic gestures. Static gestures are divided into two: Number system and gestures.

Table II. Confusion matrix for number ISL hand gestures

Number ISL Hand Gesture	Zero	One	Two	Three	Four	Five	Accuracy (%)
Zero	10	0	0	0	0	0	100
One	0	8	2	0	0	0	80
Two	0	0	10	0	0	0	100
Three	0	0	0	10	0	0	100
Four	0	0	0	0	10	0	100
Five	0	0	0	0	0	10	100
TAR	1	0.8	1	1	1	1	**96.66%**

Accuracy for the developed system is evaluated by creating confusion matrix for the gestures. True acceptance ratio (TAR) is calculated for gestures. By using confusing matrix accuracy for the individual gesture and total accuracy of the system is shown in following tables. Table II indicates confusion matrix for 0~5 number ISL hand gestures. Table III shows confusion matrix for ISL gestures victory, good, stop, candle and okandMango.

Table III. Confusion matrix for ISL hand gestures

ISL Hand Gestures	Victory	Good	Stop	Candle	Ok	Accuracy (%)
Victory	10	0	0	0	0	100
Good	0	10	0	0	0	100
Stop	0	0	10	0	0	100
Candle	0	0	0	7	3	70
Ok	0	0	0	1	9	90
TAR	1	1	1	0.7	0.9	**92%**

CONCLUSION

The main contribution is addition of background subtraction algorithm and new parameter to classify gesture with optical flow for the ISLgesturerecognition. Robustness and accuracy of the system is boosted by background subtraction algorithm. Parameters used to define the gestures are easily computed by knowing hand convex contour of hand. System is nurtured intelligently and gives warning message for critically wild luminance condition and depth of signers hand from camera if placed far away or far near. "Add gesture" tab allow user to trainon the fly to add new gesture into the system at any time. Addition of the "Actual count" tab avoids wrong training for the gesture. Use of single Naïve

Bayes classifier allowed us to train a system with multiple classes with an ease. The system is validated using five different signers. The accuracy for the ISLnumber hand gestures the 98.3 %, for ISL gesture's accuracy is 92%.To get correct optical flow signer is constrained to move hand in correct direction. This can lead wrong results this can be overcome by developing more robust algorithm.

REFERENCES

[1] Annual report 2016-17 on Indian census, By government of India; ministry of social justice and empowerment; department of disability affairs, 2016.

[2] S. Oniga, A. Tisan, D. Mic, A. Buchman, and A. Vida-Ratiu, 2007 , Hand postures recognition system using artificial neural networks implemented in FPGA, ISSE 2007 - 30th Int. Spring Semin. Electron. Technol. 2007; Conf. Proc. Emerg. Technol. Electron. Packag., pp. 507–512.

[3] V. I. Pavlovic, R. Sharma, and T. S. Huang, 1997, Visual interpretation of hand gestures for human-computer interaction: a review, IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 677–695.

[4] R. Akmeliawati, F. Dadgostar, S. Demidenko, N. Gamage, Y. C. Kuang, C. Messom, M. Ooi, A. Sarrafzadeh, and G. Sengupta,2009, Towards real-time sign language analysis via markerless gesture tracking, 2009 IEEE Intrumentation Meas. Technol. Conf. I2MTC 2009, no. May, pp. 1205–1210.

[5] N. H. Dardas and N. D. Georganas, 2011, Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques, IEEE Trans. Instrum. Meas., vol. 60, no. 11, pp. 3592–3607.

[6] M. R. Abid, E. M. Petriu, and E. Amjadian, 2015, Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar, IEEE Trans. Instrum. Meas., vol. 64, no. 3, pp. 596–605,.

[7] Q. Fei, X. Li, T. Wang, X. Zhang, and G. Liu, 2009, Real-time Hand Gesture Recognition System Based on Q6455 DSP Board, 2009 WRI Glob. Congr. Intell. Syst., pp. 139–144.

[8] A. Dhote and S. C. Badwaik, 2015 , Hand tracking and gesture recognition, in International Conference on Pervasive Computing (ICPC), 2015, vol. 0, no. December, pp. 1–5.

[9] K. Hawari, B. Ghazali, J. Ma, R. Xiao, and S. Aryza, 2012, An Innovative Face Detection Based on YCgCr Color Space, Phys. Procedia, vol. 25, pp. 2116–2124,.

[10] B. Yi, F. C. Harris, L. Wang, and Y. Yan, 2005 ,"Real-time natural hand gestures," Comput. Sci. Eng., vol. 7, no. 3, pp. 92–97.

[11] L. Feng, L. M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 5, pp. 879–891, 2015.

[12] Y. Wu and T. S. Huang,2002, "Nonstationary color tracking for vision-based human - Computer interaction," IEEE Trans. Neural Networks, vol. 13, no. 4, pp. 948–960.

[13] S. Agrawal, A. Jalal, and C. Bhatnagar, 2014, "Redundancy removal for isolated gesture in Indian sign language and recognition using multi-class support vector machine," Int. J. …, vol. 4, no. April 2016, pp. 23–38.

[14] P. Domingos, 2012 ,A few useful things to know about machine learning, Commun. ACM, vol. 55, no. 10, p. 78,.

[15] Y. Yao and Y. Fu, 2014, Contour model-based hand-gesture recognition using the kinect sensor, IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 11, pp. 1935–1944.