Journal of Theoretical and Applied Information Technology
10th January 2015. Vol.71 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RECOGNITION OF HUMAN ACTIVITIES FROM STILL IMAGE USING NOVEL CLASSIFIER

[1]GHAZALI SULONG ,[2]AMMAR MOHAMMEDALI
[1,2]utm-Irda Digital Media Centre (Magic-X),Faculty Of Computing,
Universiti Teknologi Malaysia,Utm Skudai 81310,
Johor, Malaysia
E-Mail:[1] Ghazali@Utmspce.Edu.My, [2]ammar_Ncc@Yahoo.Com

ABSTRACT

The quest for recognizing human activities and categorizing their features from still images using efficient and accurate classifier is never ending. This is more challenging than extracting information from video due to the absence of any prior knowledge resembling frames stream. Human Activities Recognition (HAR) refers to computer identification of specific activities to aid understanding of human behaviors in diversified applications such as surveillance cameras, security systems and automotive industry. We developed a novel model for classifier and used it in three main stages including preprocessing (foreground extraction), segmentation (background subtraction) to extract useful features from object and sort out these features by the classifier (classification). The model is further simulated using MATLAB programming. Our new classifier generates slightly different results for still image based on dataset INRIA and KTH for 780 images of (64*128) pixels format obtained from literature. The recognition rate of 86.2% for five activities such as running, walking, jumping, standing and sitting achieved by us is highly promising compared to the existing one of 85% over last decade.
Keywords: *Human Activities Recognition, Still Image, Segmentation, Features Extraction.*

1. INTRODUCTION

In the information technology era the Human Activity Recognition (HAR) play a paramount role in several computer applications and vision studies. HAR is based on still image or video recording. Recently, HAR using video has widely been carried out without much difficulty [1]-[5]. However, HAR using still images is difficult and much challenging task [6]. Recognition using video recording is implemented by extracting the picture frames following a process that compares current frame (one movement) with the next or previous frame to predict activities. Still image based recognition is tricky due to the absence of any prior knowledge regarding the features of event and thereby difficult to predict all the important fixed features from the captured image. The main goal of HAR using still images is to develop an automatic analysis scheme for collected data to precisely determine the human behavior. The appropriate implementation using suitable human computer interfaces critically depend on such image analyses useful for widespread applications from surveillance to security systems to automation.

Any system dealing with pattern recognition in still image should consist of stages such as noise reduction, segmentation, feature extraction, and classification. For noise removal many techniques used in literature like fuzzy inference system with edge detection to address the speckle noise image [7]. Also for segmentation methods Halime got good result with cellular neural network method using wavelet transform and special frequency to segment large number of images[8]. Habil Kalkan suggest method to select features and classification for learning the system based on estimating parameters within a nearest neighbor of Gaussian mixture model [9]

Lately, several accurate classifiers are developed for systems applications in the real time domain [10]. Most of these classifiers such as Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and K nearest neighbor (KNN) perform excellent pattern recognition. The high response time possessed by them is suitable for real time application. Few classifiers need large number of features to me more accurate. Generally, any system dealing with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recognition of patterns or activities consist of three major steps including preprocessing (noise reduction), segmentation and features extraction from the image object using classifier [11].

To recognize human behavior, it is necessary to understand each human movement by executing various activities. Any HAR system must be able to analyze the movement by recording a response for it. Often, recognition of human activities as shown in Figure. 1 is visualized by action and activities such as walking, running, and so on where parts of body movement become easily detectable for extracting features. However, for second hand motion activities such as smiling, crying and sadness one needs to use frequency domain (DT-CWT filter) to extract features [12]. Currently, image processing is widely used for medical and clinical diagnosis to detect disease by extracting features from human body with different sources including camera and scanners [13]. Biometric methods are also exploited in finger printing for user authentication [14]. Classifiers such as SVM and LDA are used to categorize features extracted from segment images. Different classifiers produce unlike results in the context of machine learning. For instance, Naïve Bayesian classifier yields better result than other for security in communication [15]. Meanwhile, SVM classifier for pattern recognition works well and require less computational time but not much efficient in security field such as emailing [16].

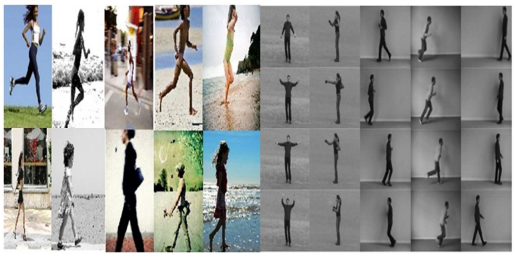

Figure 1 :Activities Within Different Dataset Using Still Images.

Yisheng et al [17] proposed a system that recognizes action automatically by tracking many still images from different cameras positioned at different angles. Four actions including walking, jumping, punching and running are recognized using Reliable Inference (RI) classifier. RI possesses two features such as MHI Motion History Image and sequence of matching frames. Juan et al [18] experimented with a combined system consisting of HAR using video and still image. In their investigation, HAR based video focused on the spatial temporal shifts and the other one with still image concentrated on static part. Despite using both static and dynamic features, the recognition rate was still low. Huimin [19] suggested a method of HAR from still images using SVM to classify the activities and achieved a recognition rate of 55%.

Fahad [20] developed a combined system of HAR using video and still image. To extract the features for HAR, local temporal dynamics of video and n-gram expressions for still image are used. He tested the method for recognizing nine activities based on decision tree classifier and obtained a HAR rate of 94% (for video) and of 70% (for still image). Two techniques such as histogram of orientation gradient and special pyramid matching are used by Bangpen [21] to recognize eight activities in still image taken from surveillance camera. He further improved the method by using scanning probe microscopy as classifier, indoor and outdoor camera until a recognition rate ~85% is achieved.

Utsav [22] used automatic software for segmentation of astrological features. He quantified the surface area and curvature of joint and surfaces to classify biomechanical geometric in three steps geometric modeling, features recognition and the development of a database structure. The author aimed to uncover local binary pattern (LBP) as a confined activity features from depth silhouettes. He recognized human activities via Hidden Markov Model (HMM) and represented the object as curve of indoor HAR for smart home using LBP [23]. Segmentation, features extraction, classification and matching techniques played important role in the pattern recognition [24]. Nae et al [25] introduced an algorithm to classify human action for the intelligent surveillance system. They used the difference between input image and the modeled background via motion information histogram. Recently, the growing concern over terrorist attacks attracted much attention to the surveillance system.

Selection of suitable classifier/model is important when dealing with still image and mixed activities (active and inactive) such as running, walking and jumping. Despite much research the accurate classifier for still image recognition is still lacking. Several classifiers for HAR exist but no one dealt with activities for still image. Our interest

Journal of Theoretical and Applied Information Technology
10th January 2015. Vol.71 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

lies in still image recognition (without focusing on response time) that do not require real time analyses particularly for activities such as reading, phoning, jogging, walking and exercising. For many activities we used INRIA and KTH dataset in our model for easy evaluation. The dataset comprised of 1580 activities in which each one contains 9 different actions. The proposed system automatically analyzes the activities using a specially designed algorithm that reduces human effort and produce instant response to the event.

2. MODELING AND SIMULATION

Schematic of the proposed system framework is illustrated in Fig.2. This consists of three major steps such as preprocessing (noise reduction and segmentation), feature extraction and classification. Because we are dealing with image as input data the noise reduction in an efficient manner is mandatory to maintain the richness of image [26]. Image segmentation is a fundamental for video and computer vision application and often used to partition an image into separate regions. Presently, several image segmentation methods are developed without much satisfactory measure in performance in which comparison among these methods become intricate [27].

We performed image segmentation following threshold method. This is a simple, powerful and yet accurate approach for segmentation of images having light objects in the dark background [28]. Our newly developed technique can efficiently and precisely classify HAR and highly suitable for still image [29]. Threshold technique is based on image space regions meaning characteristics of image. It converts a multi-level image into a binary one by choosing a proper threshold (T) and then divides the image pixels into several regions for separating the desired object from the background. Any pixel (x, y) is considered to be a part of the object whenever the intensity f(x, y) is greater than or equal to the threshold value otherwise the pixels will belong to the background. The presence of fixed camera makes the image background known. The scene starts to detect the object when it appears inside and proceed with background subtraction.

$$St(x,y) = \begin{cases} 1 & f(x,y) belongs\ to\ the\ forground\ at\ time\ t \\ 0 & herwise \end{cases} \quad (1)$$

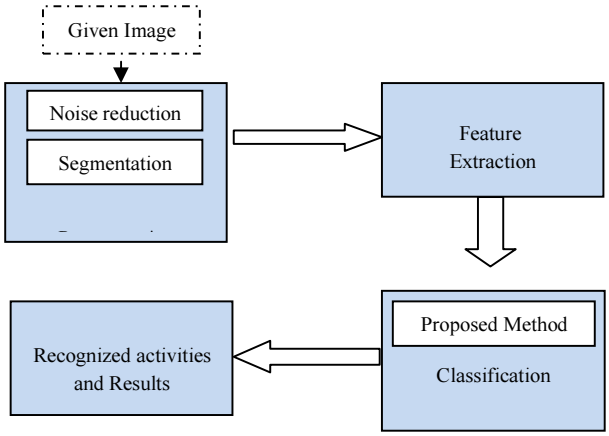

Figure 2 :Simple Framework Of The Model.

The real images are obtained via normal or night vision camera. The presence of noise in the real images severely distresses the segmentation processes and make them blurred. Therefore, the noise reduction in segmentation is compulsory [29]. The image averaging carried out for noise reduction assumes truly random distribution of noise in the image as displayed in Fig.3.

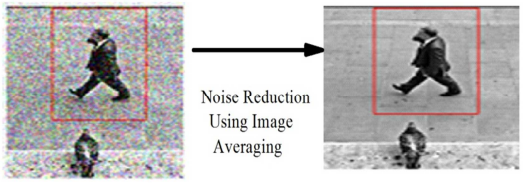

Figure 3 :Noise Reduction Using Image Averaging

The mean of N image of training data is computed. The averaging for still image is calculated by considering eight neighbor pixels in the horizontal, vertical and diagonal directions. The intensity values A(N, x, y) of pixels located at coordinates (x, y) in the averaged digital image is expressed as,

$$A(N,x,y) = \frac{1}{N}\sum_{i=1}^{N} I(i,x,y) \quad (2)$$

Segmentation with still image is more flexible because of fixed background. The object background is simply subtracted from the foreground to get the segmented object. Superior segmentation can easily extract features in the second stage as shown in Fig.2. Features are used

Journal of Theoretical and Applied Information Technology
10th January 2015. Vol.71 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as object information and supplied to the classifier to categorize. Several features compatible with the classifier are established [30].

The main idea of HAR is to extract useful information from the image and then classify them according to their features. Two types of features are extracted from an image namely the global and local characteristics. The global features comprised of the number of blobs within object or object location. Conversely, the local features which are extracted from the object itself contain length, size, skeleton and histogram of oriented gradient as shown in Fig.4. The present work attributed both the local and the global features of still image. Fig.4 (a) considers the most important parts of the human body (arms and legs) to take all features related with pixels direction depending on 8-neighborhood and allows more variation among them to set their direction as displayed in Fig.4 (b). Fig.4 (c) represents the histogram corresponding to those pixel arrays.

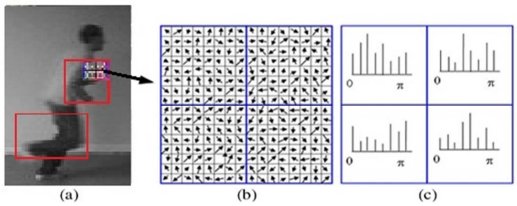

Figure 4:Procedures For Extracting Features From An Object.

Classification being the prime concept in HAR our aim is to establish a novel classifier which removes the disadvantages present in the existing one and then classify activities relating to feature distribution to be convergent. Features are divided into two categories by distinguishing them for one class and collectively combined in one group. We vertically separate classes into groups to achieve two categories (Right and Left). Each group is then horizontally divided into two equal parts (Up and Down) to locate the presence of any in-homogenous features. The separation process continued deeply until similarity features are reached. Finally, the features are tracked and stored with a binary tree as depicted in Figure 5.

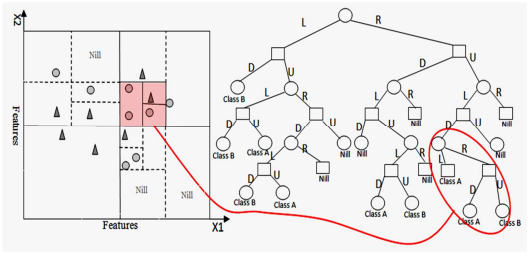

Figure 5: The Binary Storage Tree With The Novel Classifier, With R, L: Vertical Partitioning (Right And Left) U, D: Horizontal Partitioning (Up And Down) Nill: Partition With No Feature, Class A: Features Belong Class A, Class B: Features Belong Class B, ○ One Feature Of Class A And ∆ One Feature Of Class B.

The data features are stored in the terminal as binary tree for allocation and yields,

$$b_n^{(k)} = \frac{1}{n}\begin{bmatrix} n \\ 2k-1 \end{bmatrix}\begin{bmatrix} 2k \\ k \end{bmatrix} 2^{n-2k} \qquad (3)$$

where b is binary tree, n are nodes and k are leaves in terminal nodes. All the features located in the leaves can be clustered hieratically from down to top following the classification. Each cluster represents class of one activity with its inherent features.

Generally, any training example possess a validation set (x_k,y_k) and a training set $D_k=\{(x_1,y_1), \ldots,(x_{k-1}, y_{k-1}),(x_{k+1},y_{k+1}),\ldots, (x_n,y_n)\}$ with k=1, 2, 3,…, n. It predicts the class label $\hat{y}_k$ for a particular x_k using binary tree allocation $pr(y1|x1, D)$ while given data set N the data point from class k with $(x) = \frac{K}{NV}$, V is the number of features extracted.

The algorithm for the entire system is coded as:

Input image I
Output classes A, B
Begin
Read image I with gray level [1,…,L];
Apply noise reduction ……..eq(1);
Apply segmentation ……eq(2);
Extract features;
Classify the activity:
For i=1 →N do
If mixed features do divide vertically (R&L partition)
Else set class
Store in tree ……eq(3)

Journal of Theoretical and Applied Information Technology
10th January 2015. Vol.71 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

If mixed features do divide horizontally (U&D partition)
Else set class
Store in tree……. eq(3)
End for
Return 2 classes $\sum \propto_p f(I_i L_p)$
Compute recognition rate
End

The MALAB (version 7.6) soft-wares is used for simulating the model on an IBM PC (2 GB RAM).

3. RESULTS AND DISCUSSION

Our method is applied to three challenging datasets obtained from the public domain. These datasets are INRIA for 5 activities with four different actions, KTH for 5 activities given by gray image with four different movements and Willow-action in the same domain. The proposed technique aimed to recognize five main daily activities such as running, walking, jogging, boxing, and waving Fig.1. Furthermore, the training set consisting of these five different kinds of activities tried to find the action with one side (side view in right or left) without taking into consideration front and rear view. Undoubtedly, this subset of activities is chosen because of their most visually identifiable action from single images. The achieved overall accuracy rate of the supervised classification on this classifier is 86.2% which is excellent.

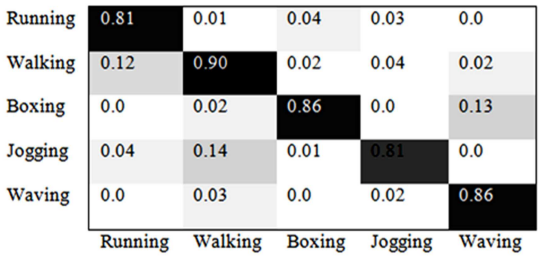

Figure 6 :Confusion Matrix Showing Recognition Rate For Five Activities.

The results for the confusion matrix with supervised classification are furnished in Fig.6. Interestingly, for activities such as walking, running and jogging the system is capable of recognizing the features extracted from two parts of the body (arm and leg) and a high recognition rate is accomplished. In contrast, the activities such as boxing and waving achieves relatively lower recognition rate because the features are extracted from one part of the body (arm).

Fig.7 represents the features dependent recognition rate for various activities. Classification of such activities critically depends on the number and character of features supported into the classifier. The classifier continues to take the same action and remain static until a fixed number of features are reached. The recognition rate is found to increase with the number of features until it reached to a critical point (seven features). Furthermore, the recognition rate remained the same unless the number of images in training set is increased.

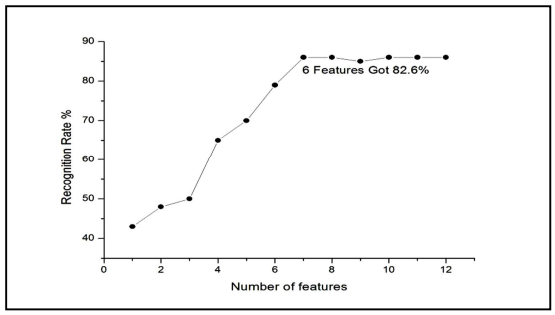

Figure 7: Recognition Rate Versus Features Obtained From The Proposed Classifier.

The only limitation of the proposed method is its unsuitability for online recognition associated with computational time which is bit longer than others. Conversely, the classifier is very accurate, efficient and highly satisfactory for still image features extraction. The recognition rate (86.2%) obtained by us is much better compare to the existing one. The recognition rate is obtained by dividing the number of times the classifier achieves correct features (success) over the number of tested images (trial) given by,

$$Recognition\ rate = \frac{Number\ of\ correct\ classified}{Number\ of\ given\ activities} \times 100\%$$

The training data for given images using dataset INRIA and KTH for five activities in the present classifier produced notable features when compared and summarized in Table 1.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Classifier	Dataset Used	Recognition Rate (%)	No. of Activities
LSVM [14]	Weizmann & KTH	70.5	7
Multi SVM+HOG [31]	Willow-action	61.07	6
SVM [32]	Willow-action	60.66	5
BSVM+LDA [33]	Collected & INRIA	85.1	6
Proposed method	INRIA, KTH & Willow-action	86.2	5

Table 1: Comparison Of Recognition Rate Of The Proposed Method With Others.

We evaluate the accuracy of our method with five activities using three different datasets because some activities are available in a particular dataset and some belong to another. Training set with INRIA and Willow dataset are found to be much more computationally time consuming than with KTH because each image is changed to gray scale followed by noise reduction, segmentation and so on. Meanwhile, training with INRIA dataset consisting of 780 images of (64 × 128) pixels format for each activity takes longer time to learn the machine.

4. CONCLUSION

A new model of classifier is developed for HAR and simulated to extract useful features from different activities using still image. The model comprised of three main steps such as preprocessing, segmentation and classification. Using existing public domain datasets we achieve the highest recognition rate of 86.2% for five activities including running, walking, jumping, standing and sitting. This value is much higher than previously reported one although our method is computationally bit time expensive. The excellent features of our result suggest that this simple classifier is potential for still image processing and information extraction.

ACKNOWLEDGEMENTS

Ammar is grateful to the Ministry of Science and Technology, Iraq for the study leave and University Technology Malaysia for technical assistance.

REFERENCES

[1] Abdel-Badeeh M, Adel A, and Usama A. , "A Vertex Chain Code Approach for Image Recognition", *ICGST-GVIP Journal* ,vol. 5, 2005.

[2] Bandera A, Urdiales C, and Sandoval F. "2D object recognition based on curvature functions obtained", *Pattern Recognition Letters,* vol. 20, 1999,pp. 49-55.

[3] Youtian D, Feng C, Wenli X, and Weidong Z, "Activity recognition through multi-scale motion detail analysis", *Science Direct,* vol. 71,2007,pp.3561–3574.

[4] Chunhui and Jitendra M, "Multi-Component Models for Object Detection", *Computer Vision Springer*, vol. 7575,13 October 2012, pp. 445-458.

[5] Chun Z, and Weihua S, "Motion- and location-based online human daily activity recognition",*Science Direct Pervasive and Mobile Computing*,vol.7, 2010,pp.256–269.

[6] Christian T, and Hlavac, V,"Pose primitive based human action recognition in videos or still images" *IEEE Computer Society Press* ,vol. 41,28 June 2008,pp. 1-8.

[7] Mehmet A, Alper B, and Mehmet E,"A novel fuzzy filter for speckle noise removal", *Turk J Elec Eng & Comp Sci* ,vol.10,2013,pp. 10-24.

[8] Halime B. and Yüksel Ö,"A new method for segmentation of microscopic images on activated sludge" *Turk J Elec Eng & Comp Sci* ,vol. 10,2013,pp.7-9.

[9] Habil K,"Online feature selection and classification with incomplete data",*Turk J Elec Eng & Comp Sci* ,vol.10,2013,pp.171-181.

[10] Pallabi P and Bhavani T," Face Recognition using Multiple Classifiers" *IEEE* , November 2006, pp. 179 – 186.

[11] Jamie S, Toby S and Alex K.," Real-Time Human Pose Recognition in Parts from Single Depth Images", *Communications of the ACM* , vol. 65,2013,pp.116-124.

[12] K. Jaya P, and R.S. Rajesh,"A Hybrid Face Recognition Approach Using Local Fusion Of Complex Dual-tree Wavelet Coefficients And Ridgelet Transform",*ICTACT journal* ,vol. 1,2011,pp.186 – 191.

[13] S. Praveenkumar," Feature Extraction OF Retinal Image For Diagnosis Of Abnormal Eyes", *ICTACT Journal* ,vol. 1,2011,pp.0976-0985.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[14] S. Malathi and C. Meena," Fingerprint Matching Based On Pore Centroids", *ICTACT Journal* ,vol. 1,2011,pp. 220 - 223.

[15] T. Hamsapriya, D. Karthika R and M. Raja C.," Spam Classification Based On Supervised Learning Using Machine Learning Teqnichues", *ICTACT Journal* ,vol.1,2011,pp.220-223.

[16] S.Aarthy and Ms.A.K.Ilavarasi," Improving the Coding Efficiency using Support Vector Machine Classifier Code Prediction", *IJRIT Journal*,vol.1, 2013,pp. 284-290.

[17] Yisheng and Jim D. Human," Activity Recognition for Synthesis", *http://cse.osu.edu/~jwdavis/Publications/Irchsv06.pdf.*2006 (retrieved on January 10, 2013).

[18] Niebles J and Li F," A Hierarchical Model of Shape and Appearance for Human Action Classification" , *IEEE* , 17-22 June 2007, pp.1-8.

[19] Huimin Q and Zhiquan W," Recognition of human activities using SVM multi-class classifier" *science direct* ,vol. 31,2009,pp.100–111.

[20] Christian T and Hlavac, V.," Pose primitive based human action recognition in videos or still images",28 June 2008,pp. 1-8.

[21] Yao B and Fei-Fei L.," Recognizing Human-Object Interactions in Still Images by Modeling the Mutual Context of Objects and Human Poses", *IEEE* ,vol. 34,2012,pp.691-703.

[22] Utsav S. and Anshuman R.," Advances in Geometric Modeling and Feature Extraction on Pots", *Arizona State University*, 2012.

[23] Md Zia U, Deok-Hwan K, Jeong T and Tae-Seong K.," An Indoor Human Activity Recognition System for Smart Home Using Local Binary Pattern Features with Hidden Markov Models", *SAGE Journal* ,vol. 22,2012,pp. 289-298.

[24] Ratnashil N., Dr. Nitin A., and Mahendra S.,"A Survey on Recognition of Devnagari Script", *IJCAIT Journal* ,vol.2,2013,pp.22-26.

[25] Nae J, and Teuk-Seob S.," Human Action Classification and Unusual Action Recognition Algorithm for Intelligent Surveillance System",*Springer IT Convergence and Security,* ,vol. 215,2012,pp. 797-804.

[26] Luisier, F. and Blu, T.," Image Denoising in Mixed Poisson Gaussian Noise", *IEEE Transactions on Image Processing*,vol.20,2012,pp.696-708.

[27] Hsieh T, Liu Y, Liao C, Xiao F, Chiang I and Wong J. ,"Automatic segmentation of meningioma from non-contrasted brain MRI integrating fuzzy clustering and region growing", *NCBI journal* ,vol. 10,2011,pp.1472-1482.

[28] Yuan W, Dejian W . ,Gang Zhang and Jun Wang,"Estimating nitrogen status of rice using the image segmentation of G-R thresholding method", *Science direct* ,vol. 149,2013,pp. 33–39.

[29] Kathy C, Min J, Bryan S, and Jun L. ,"MultiMedia Modeling", *springer*, 2014, pp. 104-115.

[30] Dopido I , Villa A , Plaza A, and Gamba P. ,"A Quantitative and Comparative Assessment of Unmixing-Based Feature Extraction Techniques for Hyperspectral Image Classification", *IEEE Journal* ,vol.5,2012,pp. 421 – 435.

[31] Weilong Y , Yang W , and Greg M. ,"Recognizing Human Actions from Still Images with Latent Poses", *cite seer journal* ,vol.2010;2.

[32] Vincent D, Josef S and Ivan L.," Learning person-object interactions for action recognition in still images",*NIPS Proceedings* ,vol.24, 2011,pp.190-199.

[33] Ikizler N, Cinbis R, Pehlivan S and Duygulu P.," Recognizing Actions from Still Images", *Tampa, FL: IEEE*, 11 December 2008,pp.1-4.