Hand Gesture Recognition system for Real-Time Application

M.Murugeswari[1] ,S.Veluchamy[2]
[1]PG Scholar, [2]Assistant Professor,
[1,2]Communication Systems,Regional office-Madurai, Anna University,Tamil Nadu, India
[1]murugeswari.mms@gmail.com,[2]pvs1834@gmail.com

Abstract— ***In*** **recent years, several researches are being done to improve the means by which human to machine interaction. With the development of input devices like keyboard, mouse and pen are not sufficient due to this limitation direct use of hand gesture as an input device to provide natural human to machine interaction. The objective of this paper is to implement the vision based hand gesture recognition system to control the movement of robot. We can use of Scale invariant feature transform (SIFT) for extract the keypoint from the gesture image capture by single sensing device. Space incompatibility of SIFT keypoint causes bag of feature approach was introduced. Then use the vector quantization will map the keypoint extracted from SIFT into unified dimensional histogram vector after the K-mean clustering. The histogram vectors as an input to multiclass SVM classifier for recognize the gesture. Generate the grammar apply to the robot to control the movements (Left, Right, Straight ward, Backward, stop) of robot.**

Keywords— Bag-of-features, Human to machine interaction, K-mean, scale invariant feature transform (SIFT), support vector machine (SVM).

I. INTRODUCTION

In the present days computer and computerized devices are important element in our society. Gestures are used to covey the meaningful information or interact to the surrounding environment through body motion like finger, head, arms, face, and hands. Multimodal gestures such as hand and arm gesture, head and face gestures and body gesture also used to control the application. The meaning of the gesture vary depends on the situation typically, spatial information (where it occurs), pathic information(path it takes), symbolic information (the sign it makes), affective information (its emotional quality).For example we say “stop,” either we can use the gesture such as raise the hand with palm facing forward or waving both hands over the head. It may be static are dynamic. Gestures used in day today when speaking on the telephone at the same time to communicate to others .Whereas hand gesture is defined as dynamic movement of hand which is referred to a sequence of hand postures connected by continuous motion like saying good bye. Gesture recognition is mainly used in sign language for hearing impaired people, distance learning/tele-teaching assistance, video surveillance and monitoring, remote control, guiding the robots. Various tools used for gesture recognition which is based on computer vision, pattern recognition and statistical modeling. Traditionally, we use the keyboard, mouse, joysticks as an input device to provide human to computer interaction (HCI).

Robots used in many real time applications. However, the robots have two problems. First, it is usually not easy for many people to operate them because the operations are often difficult and confusing. Second, safety of the robots while operating in the real environment. In this work, we focused on solving the robot operation problem. In order to solve the problem, we thought it is necessary and desirable to construct an easy operation system for everyone. Therefore, we proposed a robot operation using hand gestures. In recent days to provide human to computer by using an important technology vision based gesture recognition system. Two methods are used to provide interaction from human to computer which is data gloves based approach and vision based approach. Firstly, use of sensor devices for collecting hand and finger motion and hand motion can be converted into electrical signal. This approach needed more sensors for easily collecting the hand configuration but it is more expensive and also inconvenient to wear the gloves. In vision based approach using a camera to collect the hand motions for recognizing the hand gesture without use any extra devices for collecting hand configuration. Vision based gesture recognition is divided into two categories firstly 3-D hand model based approach can use the volumetric or skeletal model which is used in computer animation industry. A disadvantage of this approach is increase the computational complexity linearly with number of cameras used to collect the hand gesture. Secondly, use the appearance based approach utilized the skin colored region in the images. This approach is also have some disadvantages, that is no any other skin color objects are exist in the background and only take the hand images at proper lighting condition. However, vision based hand tracking and gesture recognition is a challenging problem due to complexity of hand gesture and high degree of freedom involved in human hand. Human to machine interaction by using hand gesture communication successfully fulfill their requirements in real time performance, robustness against transformation and rotation, recognition accuracy.

ISBN No. 978-1-4799-3914-5/14/$31.00 ©2014 IEEE

II. RELATED WORK

There are two categories for Vision based hand gesture recognition system which are the 3-D hand model based and appearance based methods. In 3-D model based technique a huge image database is required to deal with the entire characteristics of human hand taken by using several cameras under several views [1]. Drawback of this approach is inability to handle the image under unclear view and difficult to extract the features. In appearance based technique extract the image feature from the visual appearance of hand. It is simplest approach however this method has some shortcoming firstly it required that it sensitive to the lighting condition and also any other skin colored object exist in the image.

In [2] Lowe proposed SIFT features (keypoints) extract the features from the images. These features are invariant to scale, orientation and partially invariant to illumination change which is helpful in reliable matching between different views of same object, object recognition. Too high dimensionality of SIFT feature can be effectively solved by bag-of-feature approach [12],[13] which reduce the dimensionality of the feature space. Vector quantization technique is used in [3], using the k-mean clustering algorithm clustering the feature dimensional space vector from SIFT algorithm. This vector quantization (VQ) maps the keypoints into unified dimensional histogram vector. The clustering algorithm form the code book for each training image and the size of code book can be determine by using number of clustering nodes in the clustering process then generating Bag of word. In [4], scale space color based features are used to recognize the hand gesture however this not work well under other skin colored objects exist in the image. In [5], large set of high dimensional points can be represented by using small set of basic vectors using Eigen space technique which is not invariant to rotation, scaling and translation. The author of [6] obtained in-plane rotation invariant hand detection using Adaboost learning algorithm and SIFT features therefore, efficiency of 90.8% was achieved.

In [7], [8], Haar like features were applied for hand detection. To enhance the classification accuracy by Haar like features which is more concentrate on the particular area. Non invariant features are used when the images under scaling, rotation and translation conditions. Selections of features are important fact in [6].this can be achieved by set of Haar like feature detector. In Adaboost learning algorithm, adaptively choose the best features in each step and combine to form strong classifier. So enhance the real time performance and recognition accuracy. In [9], learning based object detection technique proposed by viola and Jones the hand detection without any restriction on the background. The detection method attains robust hand detection, but it require a large training time for obtaining the cascade classifier. With the viola- Jones detector detect the hand about 15o in-plane rotation in [10]. Even though rotation invariant hand detection can be achieved using the same way of Adaboost and treating the problem as a multiclass classification problem. Keypoints are detected by various detection methods such as SIFT in [2], Principle component analysis SIFT [11] and speed up robust features. In [2] distinctive invariant features are extracted by SIFT which is invariant to scale, rotation and partially invariant to illumination condition. SURF is fastest and has good performance as similar to SIFT but it is not stable to the rotations. In PCA-SIFT has invariant to illumination change but not stable to scale changes. Since the features were extracted in real time using SIFT algorithm.

III. PROPOSED METHOD

Our hand gesture recognition system consists of two stages: the training and testing stage. In training stage build the cluster model and SVM classifier model and in testing stage this models used to recognize the gesture. Vision based approach relies on the way human beings perceive information concerning the surroundings. However it is troublesome to implement in a efficient way. In a three-dimensional model of the human hand images are taken by one or more cameras, palm orientation and joint angle parameters are estimated. These parameters are then used for gesture classification. Second one is to capture the image using a camera then extract some feature and those features are used as input in a classification algorithm for classification.

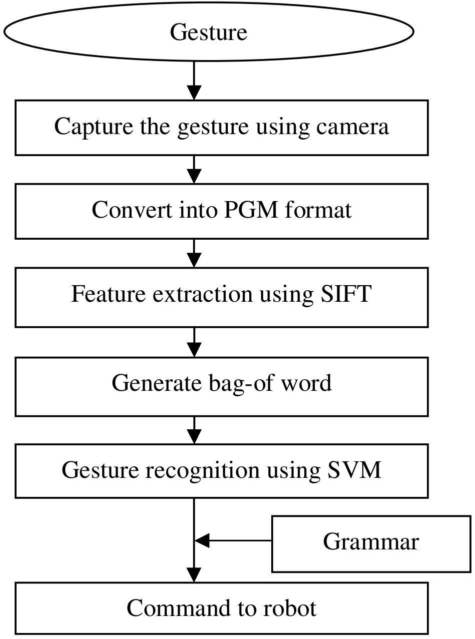
Gesture
Capture the gesture using camera
Convert into PGM format
Feature extraction using SIFT
Generate bag-of word
Gesture recognition using SVM
Grammar
Command to robot

Fig. 1. Work flow of Gesture recognition

In proposed method uses the second method specified above. SIFT algorithm [2], for feature extraction from the training image. The following stages are present the main contribution of this paper.

1) We proposes the SIFT for distinctive feature extraction from the gesture image which perform the reliable matching in testing images used for accurate gesture recognition.
2) Using Bag of feature model and SVM classifier for accurate gesture recognition and achieve real time performance.
3) To build a grammar that generates the command to control the robot even under different scale of gesture image and also transition among gesture.

Work flow of proposed method as shown in Fig. 1. Gesture image converted into gray scale (Portable Gray Map) image format then applies the feature extraction. Methodologies used for processing the gesture images will be discussed in the following sections.

A. *Features Extraction Using Scale Invariant Feature Transform (SIFT):*

The SIFT select the features are invariant to image scaling and rotation, and partially invariant to change in lighting condition and distinctive features for perform the reliable matching between the large number of image features in the database. These features are well localized along with frequency and spatial domain. SIFT algorithm having four major stages Scale space detection, keypoints localization and Orientation assignment and keypoint descriptor.

1) Scale space extrema detection:

Original image I(x, y) convolved with gaussian function varying in width i.e., different scale value of gaussian function to construct the gaussian function (scale space) of image. In this stage identify the potential interest of points over all scale and image locations. Scale space of the image is defined as a function as follows,

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma) \tag{1}$$

Where, * convolution operation

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2 - y^2\right) / 2\sigma^2} \tag{2}$$

Two nearby scales are separated by the factor 'k,' are subtracted to get Difference of gaussian images which is represented as,

$$DOG\,(x, y, \sigma, \sigma') = [G(x, y, \sigma) - G(x, y, \sigma')] * I(x, y) \tag{3}$$

Select the octave of scale space into s of intervals so K= 2^ (1/s). Suppose s is equal to 2 we produce S+3 images for each octave. So the efficiency of algorithm increased at low scale value of image.

2) Keypoint localization:

Each Sample point calculated from is compared to its nearest neighbors in the same image and the scale above and below the current image which this point is either maximum or minimum compared to the neighbors. So every time each keypoint is compared to its 26 neighbors in 3 x 3 regions. Improve the localization accuracy by using Taylor expansion. Most of the sample points are eliminated by re-iterated gaussian filtering in this stage. These feature points are not stable under noise and also irrelevant to the image description in final stage. The keypoints in the low contrast and edges are discarded. So the keypoints are invariant to the scale.

3) Orientation Assignment:

Calculate gradient orientation histogram for each keypoints in orientation assignment. Using the pixel difference compute the gradient magnitude and orientation as follows,

$$m(x,y) = \sqrt{\left(L(x+1, y) - L(x-1, y)\right)^2} + \sqrt{\left(L(x, y+1) - L(x, y-1)\right)^2} \tag{4}$$

$$\theta(x, y) = \arctan\frac{L(x, y-1) - L(x, y+1)}{L(x+1, y) - L(x-1, y)} \tag{5}$$

Each pixels are weighted by gaussian window size of the window is 1.5 times greater than the keypoints scale value. All the point which is nearer to the keypoints orientation direction is calculated using above equation. Compute the orientation histogram for each keypoints and 80% of points represent the directions as a keypoint direction. The keypoints are invariant to the orientation.

4) Keypoint Descriptor:

The keypoint descriptor is computed by calculating gradient magnitude and orientation of all the keypoint around the keypoint location. The orientation histograms are relative to the keypoint orientation. The histogram contains 8 bin and compute array of 4 x 4 histogram around the keypoint. The keypoint vector consist 128 dimensional feature vectors. So the features are invariant to the illumination. The keypoint descriptor is shown in Fig. 2. The length of arrow is the total sum of keypoints magnitude near the feature point direction within the region.

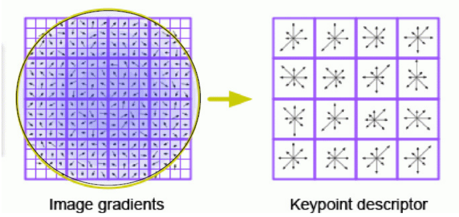

Fig. 2. Keypoint Descriptor

B. *Generating Bag-of-feature:*

Generation of bag of feature include following steps, extract the features of hand gesture image Keypoint descriptor in Sift algorithm convert the hand gesture image into 128 dimensional feature vectors [2]. Secondly, learn visual vocabulary. We choose the K-mean clustering algorithm for making the codebook (Visual vocabulary or Cluster) shown in Fig. 3.Initially we choose the centroids random in nature. Assign each keypoint vector into nearest centroid. Compute the mean of all points in keypoint vector and recomputed the cluster center in iterate manner [14], [16]. So we minimize the sum of Euclidean distance between the points to the nearest cluster center. Choose the vocabulary size based on the structure of the image data. If the vocabulary size is too small, visual words not representative for all Key point vectors. If vocabulary size is too large, quantization artifacts and over fitting problem can occur. Third, quantize features using visual vocabulary. Vector quantization (VQ) each vector is mapped to the code vector of a codebook and to find reproduction vector which reduces the codebook dimensionality.

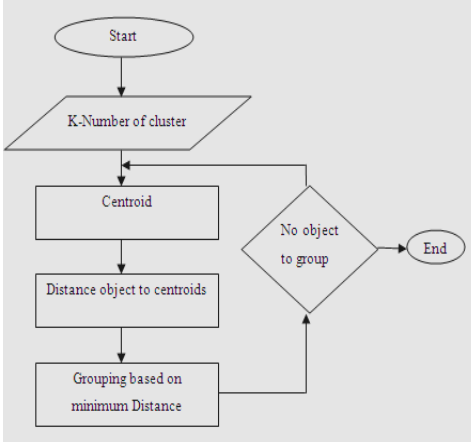
Start
K-Number of cluster
Centroid
Distance object to centroids
Grouping based on minimum Distance
No object to group
End

Fig. 3. K-mean clustering algorithm

Finally, we encode each feature from gesture training image to obtain the histogram (i.e. represent images by using frequencies of visual word) [16].

C. Multi-class support vector Machine (SVM) classifier:

It follows supervised learning method used for classification and regression. A hyperplane or set of hyperplane in high dimensional space is provided by SVM shown in Fig. 4. In SVM uses the maximum margin hyperplane provides maximum separation between the classes [19]. A good separation between different classes of feature vectors can be achieved. In Multi-class SVM for Binary classifications which convert the multiclass problem in to two class problems. Two common methods in adaptation which are 1A1 and 1AA [20].In proposed method include one- against-one approach. In 1A1 approach build a machine for each classes (i.e. m (m-1)/2). Two class classifiers are created). When apply a training image train the classifier using a sample of classes as a positive example and second class as a negative example. Each classifier gives a vote to the particular class and maximum vote are earn by particular class. Winning class is used for gesture recognition in testing stage.

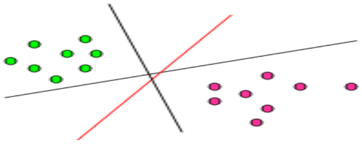

Fig. 4. Separation of feature vectors by optimal hyperplane

IV. EXPERIMENTAL RESULTS

We tested five hand gestures. Cambridge hand gesture data set consists of 900 images which classified as 9 classes the dataset having gesture images taken under different motions leftward and rightward and different illumination condition each class contain 100 images.

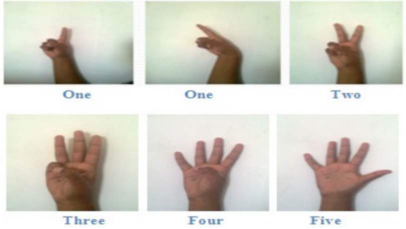

Fig. 5. Sample images from Database

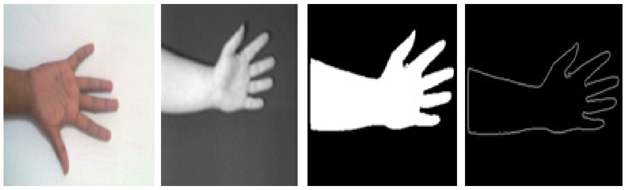

Fig. 6. Preprocessed images: a) Input image b) Image in PGM format c) Binary image d) Contour Detected image

Create database for training stage. Database is any particular location in system memory. It contains image taken under different scale i.e. image taken at nearer to the camera and larger distance from the camera. To take images using camera is also used for processing. The database contains number of training images which increase the robustness of the algorithm. Fig. 5 shows the some sample images from the database. The image processing time is too important in training stage as well as training stage. We can reduce image processing time by reducing the number of keypoints. This is achieved by reducing image resolution and converting the image into PGM format show in Fig. 6b.Using Otsu’s method convert gray scale images into binary image. In testing stage skin color based method for the hand detection. Contour of the hand gesture is used detecting the hand in real time processing and eliminating other skin colored object exist in the image shown in Fig. 6.

SIFT algorithm used to extract the features from the training image. There is minimum number of features used to present the object. Too few features are may not used to present the object clearly. If too many features are used it require the more processing time and more memory space. The hand is away from the camera, to extract small amount of keypoints from that image. If the hand is nearer to the camera, we extract large amount of keypoints from the image. Scale invariant keypoints are extracted and also these features are invariant to illumination of light and invariant to translation. The input image is convolved with the gaussian function.

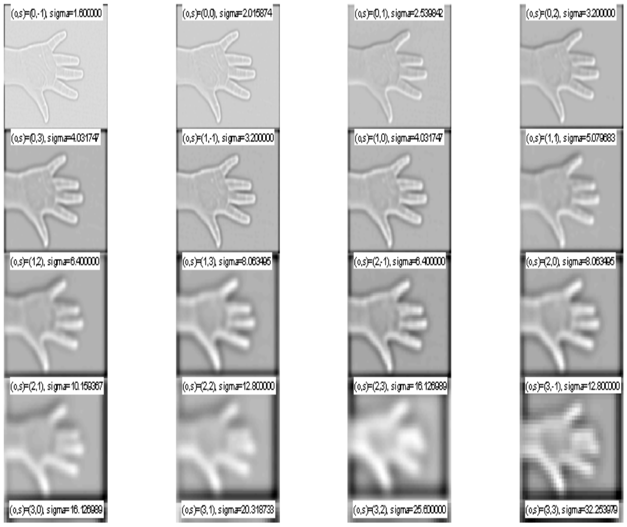
(o,s)=(0,-1), sigma=1.600000
(o,s)=(0,0), sigma=2.015874
(o,s)=(0,1), sigma=2.539842
(o,s)=(0,2), sigma=3.200000
(o,s)=(0,3), sigma=4.031747
(o,s)=(1,-1), sigma=3.200000
(o,s)=(1,0), sigma=4.031747
(o,s)=(1,1), sigma=5.079683
(o,s)=(1,2), sigma=6.400000
(o,s)=(1,3), sigma=8.063495
(o,s)=(2,-1), sigma=6.400000
(o,s)=(2,0), sigma=8.063495
(o,s)=(2,1), sigma=10.159367
(o,s)=(2,2), sigma=12.800000
(o,s)=(2,3), sigma=16.126989
(o,s)=(3,-1), sigma=12.800000
(o,s)=(3,0), sigma=16.126989
(o,s)=(3,1), sigma=20.318733
(o,s)=(3,2), sigma=25.600000
(o,s)=(3,3), sigma=32.253979

Fig. 7. Difference of Gaussian Images

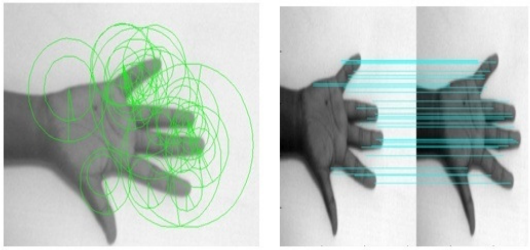

Fig. 8. Keypoint Detection and matching by SIFT

TABLE I. Performance Comparison of Various Feature Detection methods

Method	Time	Scale	Rotation
SIFT	Good	Best	Best
PCA-SIFT	Good	Common	Good
SURF	Best	Good	Common

The difference of gaussian images are computed and shown in Fig. 7. The keypoints are placed along the edges and affected by noise are eliminated by using SIFT. Compute orientation histogram for the image orientation assignment. So we detected the features are invariant to scale and orientation. The SIFT select the distinctive features which enable the correct matches with the large database. For image matching the feature are extracted from the training image and stored in the database. In testing stage each time apply a new image and compute the match with the features in the database. Matching the testing image with the training image is shown in Fig. 8.

In real time testing process as we took 4 classes of gesture image on 12 mega pixel digital camera. Using SIFT extracted feature are represented above. We assign a first 2 classes of images into codeword 1 and remaining two are in codeword 2.

Then, put the both two code words into common codeword. This is the codeword vector after generating bag of word. As we applied the n x m size image that converted into unified dimensional vector at the end of generating Bag of word. Apply this to multiclass Svm Classifier. That convert m-class problem into two class problem. In proposed method 4 classes of gesture are used (4x3)/2=6 binary SVM for gesture classification use max win technique for gesture recognition.

TABLE II. Experimental Results for Gesture Recognition

Gesture	No. of test images	Recognized images
One	50	49
Two	50	48
Three	50	46
Four	50	47
Five	50	46

TABLE III. Comparison of classification rate of different classifiers

Classifier	System efficiency (%)
HMM	79
ANN	86
SVM	97

HMM-Hidden Markov Model, ANN-Artificial Neural Network
SVM – Support Vector Machine

V. CONCLUSION

In this paper, we provide interactive communication between human to computer using vision based approach. In this approach, only uses the single sensing device to capture the hand gesture no need extra devices that reduce the cost of the process. We used the SIFT algorithm to extract the keypoints (vectors) from each gesture image. The number of keypoints decreases when the hand gets away from the camera and increases when the hand comes closer to the camera, because the area of the hand increases. We captured the hand gesture image for different people with different scales, orientation and illumination conditions. SIFT are invariant to scale, orientation, and partially to illumination changes. To extracting distinctive invariant features from images that can be used to perform reliable matching using SIFT. Then apply the vector quantization technique which clusters the keypoint descriptors in their feature space into large number of clusters. Encode each keypoint by the index of the cluster (code vector) to which it belongs. Generate bag of word to apply this to multiclass SVM classifier achieve the recognition accuracy up to 97 %. Recognize the hand gesture which is presented to robot. In future, the user can interact with robot and guide this by using various gestures. Based on the command received robot movement is controlled. The movement of the robot depends on the gesture identified by the system.

ACKNOWLEDGMENT

The project work was supported by Department of communication systems, Regional office of Anna University, Madurai. Authors would like to thank all faculty members for greater support and encouragement.

REFERENCES

[1] A EI-sawah, N. Georganas, and E. Petrriu, “A prototype for 3-D hand tracking and gesture estimation,” IEEE Trans. Instrum. Meas., vol. 57, no. 8,pp. 1627-1636,Aug.2008.

[2] D.G Lowe, “Distinctive image feature from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2,pp. 91-110, Nov. 204.

[3] A.Bossch, X. Munoz, and R.Martri, “Which is the best way to organize/classify images by content” Image Vis. Comput., vol. 25, no. 6,pp. 778-791, Jun. 2007.

[4] L. Bretzner, I.Laptev, and T. Lindeberg, “Hand Gesture recognition using multiscale color features, hierarchical models and particle filtering,” in Proc. Int. Conf. Autom. Face Gesture Recog., Washington, DC, May 2002.

[5] A. Argyros and M. Lourakis, “Vision based interpretation of hand gestures for remote control of a computer mouse.” in Proc. Workshop Comput. Human Interact, 2006,pp. 40-51.

[6] C.Wang and K.Wang, “Hand gesture recognition Using Adaboost with sift for human to robot interaction,” vol.370.Berlin, Germany: Springer-Verlag, 2008.

[7] A. Barczak and F.Dadgo star, “Real-time hand tracking using a set of co-operative classifiers based on Haar-like features,” Res. Lett. Inf. Math.Sci., vol. 7, pp. 29–42, 2005.

[8] Q.Chen, N.Georganas,and E. Petrriu, “Real-time Vision based hand gesture recognition using Harr-like features,” in Proc. IEEE IMTC,2007,pp. 1-6.

[9] P.Viola and M.Jones, “Robust real time object detection,” Int. J. Comput. Vis. Vol. 2. No.57,pp 137-154, 2004.

[10] M. Kolsch and M. Turk, “Analysis of rotational robustness of hand detection with a Viola-Jones detector,” in Proc. 17th ICPR, 2004, pp. 107–110.

[11] Y. Ke and R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2004, pp. II-506–II-513.

[12] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2006, pp. 2169–2178.

[13] Y. Jiang, C. Ngo, and J. Yang, “Towards optimal bag-of-features for object categorization and semantic video retrieval,” in Proc. ACM Int. Conf. Image Video Retrieval, 2007, pp. 494–501.

[14] N. Dardas, Q. Chen, N. Georganas, and E.petriu, “Hand Gesture recognition using bag-of-features and multiclass SVM,” in Proc. 9th IEEE Int. Workshop HAVE,Phoenix,AZ, Oct. 16-17, 2010,pp. 1-5.

[15] R.Gonzalez, R.Woods, and S.Eddins, Digital Image Processing Using MATLAB. Englewood Cliffs, NJ:Prentice-Hall, 2004.

[16] H.Nasser Dardas and Nicolas D. Georganas, “Real Time Hand Gesture Detection and Recognition using Bag of features and support vector machine Technique,” in IEEE Trans. On Instru. And Measurements, Vol. 60, no. 11, Nov 2011,pp 3592-3607.

[17] C.Chang and C.-J. Lin, LIBSVM: A Library for Support vector machines,2001. [online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[18] J. H. Friedman, “Another approach to polychotomous classification,” dept. Statist., Standford Univ., Stanford,CA,1997.

[19] J.Weston and C. Watkins, “Multi-Class support vector machines,” in Proc. ESANN,M. Verleysen, ED.,Brussels,Belgium,1999.

[20] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multi-class support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.