Face Recognition System Using Bag of Features And Multi-Class SVM For Robot Applications

Salah NASR
CEM Laboratory ENIS Sfax
National Engineering School of Sousse
University of Sousse, Tunisia.
nasrsalah.doc@gmail.com

Muhammad Shoaib
DIP Research Lab
Islamia College University Peshawar.

Kais BOUALLEGUE
Department of Electrical Engineering
Higher Institute of Applied Sciences
and Technology of Sousse.

Hassen MEKKI
CEM Laboratory ENIS Sfax
National Engineering School of Sousse
University of Sousse, Tunisia.

***Abstract*— Face recognition system is used for the identification and verification of a face from a video or digital image. In the first phase, Viola Jones algorithm is used to detect and crop face region automatically from image/video frame. The second phase is to recognize the face of a person, in our proposed method Bag of Word technique used to extract features from an image which uses SURF for interest point selection, the proposed technique is tested on different state of the art face recognition databases and found more accurate and fast for face recognition. Our proposed method recognizes more than two person faces, multi-class support vector machine is used to classify the face image and assign a class label based on the learning of the classifier. The algorithm for the detection and recognition of faces is implemented in MATLAB application, and achieves a high accuracy rate of 99.21, which will be tested on the mobile robot.**

***Keywords*— ; Face recognition; SURF features ; Multi-Class SVM; Classification; Robot .**

I. INTRODUCTION

Face recognition is becoming a multi-dispensary research area that has vast applications in robotics, the identification and verification of a person in a video or digital image, the applications of face recognition ranges from criminal detection from a national database to social media website, it is mostly used in security systems and also used for the identification check of a person on the internationals borders, the human faces are different from each other as like finger prints and eye irises, the national academy of science addressed that face recognition is an important subject in biometrics, in the first phase of face recognition the faces are isolated from an image with multiple faces. There are many algorithms for the automatic detection of facial, Viola Jones algorithms is one of the most popular algorithm for face detection[1]. Using the face detection algorithms faces are easily cropped from an image that contain multiple faces, many datasets doesn't require face detection i.e. AT&T face recognition dataset [14]. Among biometric techniques face recognition is not the most reliable and efficient technique, an advantage of face recognition has over other biometrics is that, it doesn't require the cooperation of the subject.

Facebook uses deep-face system in which 3D face model is employed to apply a piecewise affine transformation[2]. A nine layer deep neural network is used to derive a face, the system is trained with the largest labelled dataset of four million faces belonging to four thousand individuals, a simple classifier reaches an accuracy of 97.35%. A new feature extraction method is used for face recognition that uses the Hough transform peaks[3], these peaks are used for feature extraction and binary particle swarm optimization (BPSO) is used for selecting optimal features from the feature vector, and testing image is classified using Euclidean classifier. The Euclidean distance of a Eigen face is compared with the stored Eigen faces, the least Euclidean distance between the new Eigen face and the stored Eigen face is considered the matched face[4]. A new technique Learned Local Gabor Patterns (LLGP) is used for face representation and recognition. This technique is based on Gabor features [5]. A technique is proposed in [6] the intensity mapped unsharp masking is used to reduce the intensity of the background of a face image, the edges are improved using Laplacian of Gaussian filter and BPSO is used for optimal feature selection and Euclidean classifier is used for classification. A 2D Laplacian face method is used for feature extraction[7], this method is more accurate than one dimensional Laplacian faces and KNN is used for classification purpose. A novel feature extraction method for face images classification is used that has many advantages over other traditional feature extraction method and these features are more suitable for face

recognition, the recognition is done using KNN classifier. A new face recognition system is proposed that uses scale invariant feature transform for the extraction of facial features[8], these features are robust and fast, but are weaker and has more time complexity than SURF[9].

The face recognition system may fail on poor lighting condition, long or small hairs, beard and glasses. A number of techniques used for recognition of faces, i.e. PCA [10][11]using Eigen faces, linear discernment analysis[12]and hidden morrow model[13], faces can also be recognized using traditional methods used for features extractions and classifications. In machine learning classification is a supervised learning technique that classifies labelled data. To recognize only two faces, a binary classifier can be used for satisfactory recognition of faces, in case of recognizing more than two faces a multi-class classifier is used. The feature is a piece of information of an image that are used in machine learning for image classification, content based image retrieval and matching of images[14]. In our proposed method bag of features method is used for features extraction and multi-class SVM is used to test the query image, this system can be used for real time or offline face recognition.

The rest of this paper is organized as follows: The proposed methodology is introduced in the next section. Then, the experimental results are given in section III. Our concluding remarks are contained in the final section.

II. PROPOSED METHODOLOGY

The methodology of the paper consists of three parts (i.e. face detection, feature extraction and classification) which is described below.

A. *Face Detection*

The faces are automatically detected using viola Jones algorithm, which is one of the most used algorithms for face detection, the system is trained with face and non-face images, this technique can easily detect single and multiple faces from images and video. The face regions are automatically detected and cropped from the standard database of face recognition systems as described in Fig.1.

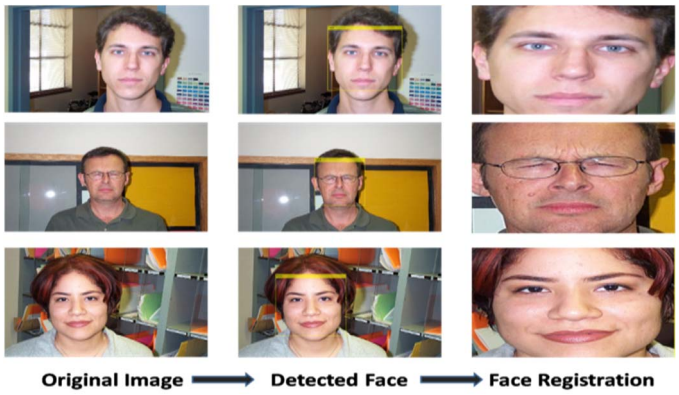

Fig 1. Automatic Face detection and cropping using viola jones algorithm .

B. *Features Extraction*

Feature extraction using BOF: Bag Of Feature is also called Bag of Words, in this method local features of training images are extracted, speeded up robust features is a local feature descriptor that is used for object recognition, registration and classification. The working mechanism of SURF features is similar to SIFT but is faster and robust than SIFT, SURF descriptors are mainly used for object and face recognition.

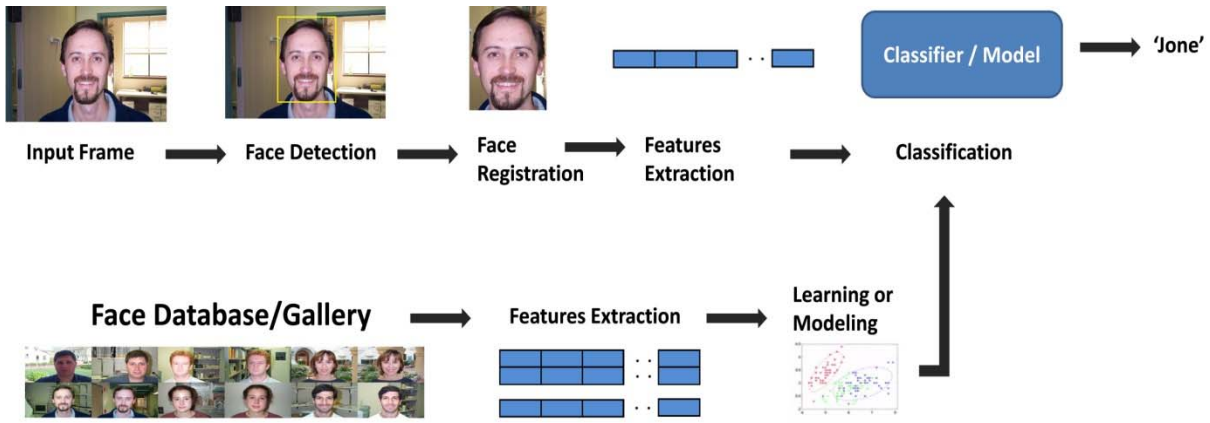

Fig. 2 Our proposed framework for face recognition system.

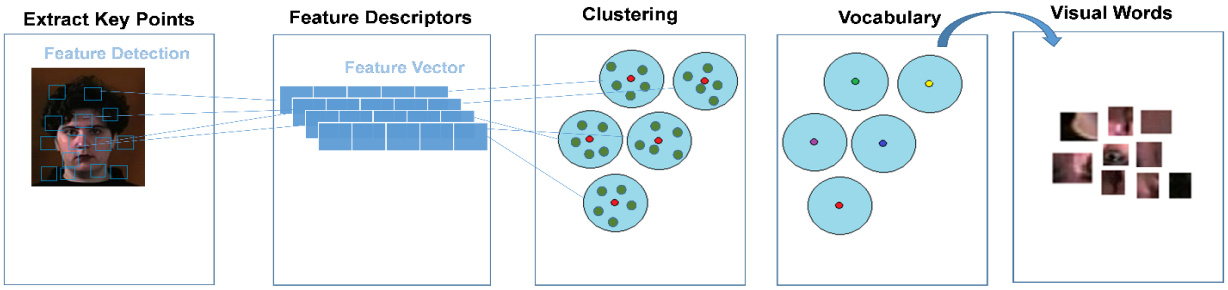

Fig 3. Create a visual vocabulary, or BoW, by extracting feature descriptors.

SURF algorithm is implemented in three phases, the first phase is points of interest selection, hessian matrix is used for finding interest points in an image, in the second phase of surf local neighbourhood descriptor is used to describe the pixel intensity distribution with the neighbour of point of interest, in the third phase different features are compared that are obtained from different images and matching pairs. The point of interest selection is done using a grid with spacing [8x8], a vector of block width 32 pixels is created, where each element of the vector corresponds to a square block from where SURF features are extracted, multiple square blocks are used to extract multi-scale features. The points of interest are stored in feature vectors and these features are quantized using K-mean clustering algorithms, by default number of clusters used by bag of feature is five hundred that can be increased or decreased. Visual words are the centers of the clusters and vocabulary is the combination of all visual words, the features in vocabulary are categorized and a visual word is used to represent the category as described in Fig.3. An encoding method is used for counting visual word occurrences in an image, a histogram of the image is created that is reduced repetition of an image.

C. Classification

As our dataset consists of more than two faces images, instead of using a binary classifier our proposed method uses multi-class Support Vector Machine to classify the face images. The multi-class SVM is trained using the error-correcting output code framework, multi-class SVM uses bag of features to encode images from the training set into histogram of visual words. These histograms are used to train the classifier; each image from the training dataset is encoded using the encoding scheme of bag of features algorithm, features from every image are detected and extracted and then features histogram of each training image is created with the help of an approximate nearest neighbour clustering algorithm. The length of the histogram is dependent on the number of visual words and a feature vector is created from the histogram.

Fig. 4 Encoding technique to encode all the training images

Support vector machine is a popular classifier that has vast application in image classification and data mining, the SVM was designed to classify only two classes using the binary SVM for multi-class classification it encounter too many problems, the binary SVM is extended to multi-class classifier that reduced the problems which were occurring in binary classifier to classify more than two classes[6]. SVM select a hyperplane between two classes of the training data that has maximum margin, many hyperplane can be drawn to separate the two classes of training data, SVM select the optimal hyperplane. The hyperplane can be defined as:

$$wx+b=0 \quad (1)$$

Where x is a point on the hyperplane, w represent normal to the hyperplane and b is the bias,

$\frac{|b|}{||w||}$ represent the perpendicular of the hyperplane to the origin. $||w||$ is the mean of the Euclidean.

Hyperplane for separating two classes in case linear separable can be defined as:

$$wxi+b \geq +1, \; yi = +1 \quad (2)$$

$$wxi+b \leq -1, \; yi = -1 \quad (3)$$

Combining equation (2) and (3)

$$yi\,(wxi+b)-1 \geq 0 \quad (4)$$

The error correcting output code framework is a multi-class model that reduces three or more classes to binary class, our proposed method uses a coding design known as one vs one, the ECOC model was introduced to combine binary problems

for the sake of solving multi-class problems, this framework is widely used in computer vision systems for face recognition, finger print recognition and handwriting recognition. The ECOC framework divides the problem into two stages i.e. encoding and decoding, ECOC is applied in two steps as presented in TABLEI:

TABLEI. ONE VS. ONE CODING SCHEME.

	Learner 1	Learner 2	Learner 3
Class 1	1	1	0
Class 2	-1	0	1
Class 3	0	-1	-1

Training of learner 1 on the observations of class 1 and class 2, according to learner 1 class 1 is positive while class 2 is the negative class, similarly the other learners are trained using the same training scheme. Suppose M is a coding design matrix with elements of mk_l and S_l is the predicated classification score for the positive class.

On the basis of training of the learner, in observation the class($\hat{k}$) assigns:

$$\hat{k} = \arg_k \min \frac{\sum_{l=1}^{L} |mk_l| g(k_l S_l)}{\sum_{l=1}^{L} |mk_l|} \quad (5)$$

III. EXPERIMENTAL RESULTS

Our proposed face recognition system is tested on some benchmarked dataset, i.e. ORL face database, frontal face dataset, and face recognition data. The grid size for feature selection is 8x8 and block width is [32, 64,96,128] pixels, for the feature quantization the number of clusters used is five hundred.

A. AT and T Face Dataset

The ORL face dataset contains images of 40 persons, this is a standard dataset for face recognition systems. The images are captured in a different time, lighting variation and different facial expression, the images of the individuals are taken from the frontal position and some with a little side movement. These images are 8 bit grayscale and are stored in PGM format, the database consist of 40 folders each 10 images per folder.

TABLEI. RESULTS OF OUR PROPOSED METHOD ON THE AT AND T FACE DATASET WITH DIFFERENT NO. OF IMAGES AND CROSS VALIDATION.

No: classes	Training Image Per class	k-size	Cross validation %		Time in Seconds		No: of features	Avg Accuracy %	
			Training	Testing	Training	Testing		Training	Testing
40	10	500	70	30	19.29	0.59	150520	100	99.8
40	10	500	30	70	10.68	0.36	64520	100	99.0
40	03	500	70	30	7.19	0.26	43000	100	93
40	03	500	30	70	5.13	0.25	21520	100	79

B. Frontal face dataset

This dataset is created by California institute of technology, the total number of face images is 450 with a resolution of 896x592 pixels, the dataset consist of face images of 27 individuals and the images are stored in JPEG format. These face images have light variation, facial expression and different backgrounds.

TABLEIII. RESULTS ON THE FRONTAL FACE DATASET, WITH DIFFERENT NO. OF IMAGES AND CROSS VALIDATION.

No: classes	Training image Per class	k-size	Cross validation %		Time in Seconds		No: of features	Accuracy %	
			Training	Testing	Training	Testing		Training	Testing
18	18	500	30	70	12.56	0.97	81985	100	91
18	18	500	70	30	28.27	0.79	205310	100	100
18	03	500	30	70	4.64	0.86	28000	100	70
18	03	500	70	30	8,07	0.81	56560	100	80

C. Face recognition data

The face recognition data consists of four directories, among the four directory the face-95 and face-96 are difficult face images to recognize.

- **Face-95 dataset:**

This dataset consist of 72 individuals face images of male and female with red background, most images of frontal face and some with minor side movement, some of these images consist of some translation. In this dataset there is lighting variation and some facial expression, the resolution of each image is 180x200 pixels and are stored in JPEG format. TABLE.IV shows the results on the face recognition data Face 95.

TABLE IV. RESULTS ON THE FACE RECOGNITION DATA, FACE-95 WITH DIFFERENT NO. OF IMAGES AND CROSS VALIDATION.

No: classes	Training image Per class	k-size	Cross validation %		Time in Seconds		No: of features	Avg Accuracy %	
			Training	Testing	Training	Testing		Training	Testing
72	20	500	30	70	40.07	0.39	364,320	100	99
72	20	500	70	30	107.00	0.95	850,080	100	100
72	03	500	30	70	9.23	0.39	60,720	100	92
72	03	500	70	30	15.31	0.40	121,440	100	91

- **Face-96 dataset:**

The face directory consist of 140 individuals face images with 196x196 pixels, the only difference in face-95 and face-96 is changed in the background and scale of the face. The background in the face-96 dataset is very complex. TABLE.V shows the results on the face recognition data face 96.

TABLE V. RESULTS ON THE FACE RECOGNITION DATA, FACE-96 WITH DIFFERENT NO. OF IMAGES AND CROSS VALIDATION.

No: classes	Training image Per class	k-size	Cross validation %		Time in Seconds		No: of features	Accuracy %	
			Training	Testing	Training	Testing		Training	Testing
33	20	500	70	30	128.74	0.88	858,000	100	99
33	20	500	30	70	37.48	0.42	330000	100	95
33	03	500	70	30	18.75	0.31	132000	100	94

33	03	500	30	70	12.09	0.24	66000	100	85

TABLE VI. COMPARISON OF SEVEN DIFFERENT APPROACHES ON ORL DATABASE[12].

Method	Top recognition rate	Running time
PCA	95.50	21.78
2DPCA	97.00	7.89
Alternative 2PCA	97.50	6.73
$2D^2$ PCA	97.75	3.84
2DLDA	98.00	6.84
Alternative 2DLDA	98.00	7.05
$2D^2$LDA	98.50	4.14
Proposed	99.21	3.24

IV. CONCLUSIONS

In our work, the face representation via bag of features for face recognition has been proposed, then linear SVM has been used for efficient classification of a face. In conventional BoW method SIFT features are used for creating dictionary, in our proposed method SIFT features are replaced with SURF features for fast and accurate features extraction. Extensive experiments demonstrate that the proposed method can achieve very good performance and outperforms most of state-of-the-art methods. we also have developed a local dataset of face images using raspberry pi camera, the system was trained with local created dataset and achieved a high accuracy of 99.21. It is also shown that, the proposed method can well handle different sizes of images as well as 1000 large number of categories, and make full use of a single image per subject for training. Moreover, it is very robust to the variations of expressions, illuminations and poses etc.

This work will be investigated to control the mobile robots like a wheelchair for disabled persons. The robot will successfully recognize the facial movement and move forward, backward, right, left and stopped when the movement of the user was detected.

REFERENCES

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Comput. Vis. Pattern Recognit., vol. 1, pp. I–511–I–518, 2001.

[2] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, "DeepFace : Closing the Gap to Human-Level Performance in Face Verification." , pp. 1701-1708,2014.

[3] R. Varun, Y. Vivekanand, K. Manikantan, and S. Ramachandran, "Face Recognition using Hough Transform based Feature Extraction," Procedia - Procedia Comput. Sci., vol. 46, no. Icict 2014, pp. 1491–1500, 2015.

[4] M. Çar and F. Özen, "INSODE 2011 A Face Recognition System Based on Eigenfaces Method," vol. 1, pp. 118–123, 2012.

[5] G. Sitaram, S. Sarkar, K. Manikantan, and S. Ramachandran, "DWT feature extraction based face recognition using intensity mapped unsharp masking and laplacian of gaussian filtering with scalar multiplier," vol. 6, pp. 475–484, 2012.

[6] S. Yang, G. Bebis, Y. Chu, and L. Zhao, "Effective Face Recognition Using Bag of Features with Additive Kernels," no. 2.2012.

[7] S. Govindarajulu and K. N. K. Reddy, "A Comparison of SIFT , PCA-SIFT and SURF," vol. 1, no. 3, pp. 53–60, 2014.

[8] C. J. Zhou, L. Wang, Q. Zhang, and X. P. Wei, "Face recognition based on PCA and logistic regression analysis," Optik (Stuttg)., vol. 125, no. 20, pp. 5916–5919, 2014.

[9] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular PCA approach," Pattern Recognit. Lett., vol. 25, no. 4, pp. 429–436, 2004.

[10] S. Noushath, G. Hemantha Kumar, and P. Shivakumara, "(2D)2LDA: An efficient approach for face recognition," Pattern Recognit., vol. 39, no. 7, pp. 1396–1400, 2006.

[11] P. Corcoran and C. Iancu, "Automatic Face Recognition System for Hidden Markov Model Techniques," Face Recognit. Vol. 2, Intech …, 2011.

[12] G. M. Foody and A. Mathur, "No Title," vol. 1343, pp. 1335–1343, 2004.

[13] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, "Learned local Gabor patterns for face representation and recognition," Signal Processing, vol. 89, no. 12, pp. 2333–2344, 2009

[14] 'The ORL Database of Faces'