# High accuracy Face Recognition system based on SIFT

**Ajitkumar Deshmukh, Pritam Sirpotdar, Juber Sheikh, Prof. Dr. M. A. Joshi**

***Abstract*— we present a robust Face Recognition (FR) algorithm based on scale invariant feature transform. As it is based on SIFT, this algorithm is also invariant to translation, rotation, scale and illumination parameters. It is also robust to noise in the image for FR. We find local descriptors from image by using SIFT, but as there are number of local descriptors in every image, we select only highly stable local descriptors. In order to reduce the computational complexity and dimension of the feature vector which contain information about the local descriptors, a detailed study of face recognition using SIFT has been done. As we are selecting only highly stable local descriptors, every local descriptor is matched accurately with high probability against large database of local descriptors from many images. The algorithm based on SIFT show the results which are very accurate and very good at discrimination of the features of different images and provide efficient recognition of the face. It is robust against the challenges like illumination, pose variation and expression change, etc. It gives recognition rate more than 92%.**

***Index Terms*—Face recognition, SIFT, similarity matching algorithm, DoG.**

## I. Introduction

Face recognition has been an important research topic amongst researchers over a past few decades. Face recognition is an important part of biometrics. Now-a-days, robust face recognition systems are in great demand to help against crime and terrorism. Until now, there are many algorithms proposed for face recognition. But every algorithm has its advantages and disadvantages. However, identification of a person in large database with high accuracy is still a big challenge for researchers. Because of the variability in the human face, the different conditions such as illumination, rotation, expression, camera view point, aging, make up and eye glasses. Often these various parameters greatly affect the performance of any face recognition system; especially when the system needs to match an input image with a large database.

In this paper, we have proposed a face recognition technique which is based on Scale Invariant Feature Transform (SIFT). Here SIFT extract local features from the input image for entire database. The next step will be similarity matching algorithm which is used to match the input image features with the features stored in trained database. Because of using similarity matching algorithm for matching the training and testing images, the accuracy of recognition goes up to 94% which makes the proposed system more robust and accurate than any other previously proposed face recognition systems. For experimental purpose, we have made use of ORL (AT&T) database which consists of 400 images of 40 different individuals. The database consists of images of these 40 individuals with different pose and expressions.

***Manuscript received June, 2015*.**

***Ajitkumar Deshmukh**, E&TC Department COEP, Pune, India, 8793232399*
***Pritam Sirpotdar**, E&TC Department COEP, Pune, India, 9421137481,*
***Juber Sheikh**, E&TC Department COEP, Pune, India, 9673118426,*
***Prof. Dr. M. A. Joshi**, Fellow IET COEP, Pune, India, 9822013631.*

## II. Review of Literature

Face recognition has been used for person identification and verification for security purpose from last two decades. But the techniques which are proposed for face recognition can be classified into two categories; first is global features extraction technique and second is local features extraction technique [4]. Global features extraction technique uses whole face image in order to extract global features. Global features include features which can be extracted by Principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA) algorithms which require the entire face image for processing [2,6,5]. These methods have common problem for face recognition such as illumination, pose variation and expression change [5]. To minimise such kind of problems, they divide face image into smaller blocks and then apply global feature extraction algorithms on the divided blocks of image [5]. Although the problem remains unsolved, they still don't achieve the desired accuracy and less computational complexity for their algorithms. In local feature extraction techniques, only those features are extracted which are considered to be important from the input face image [7, 9]. This technique extracted the geometric facial features from the input image such as mouth, eyes, brows, and chicks and the geometric ratios between them. The features extracted by local feature extraction techniques are more stable than features extracted by global feature extraction techniques. SIFT is local feature extraction technique which is used in face recognition, object tracking and recognition algorithms [7, 4, 10, 16, 1]. SIFT simply transforms data present in the face image into the key point descriptor which is going to be used as local feature of that particular image. Initially, in



ISSN: 2278 – 1323 

1999, David Lowe proposed the SIFT. It was used for object recognition after that Lowe proposed an algorithm for face recognition based on SIFT which was almost same algorithm that he used for object recognition.

## III. Related work

SIFT method implementation: the system proposed by David Lowe in 2004 to extract the features which are distinctive in nature, by using SIFT [8]. SIFT transform an image into a large collection of local feature vector which are invariant to image scale, rotation, translation, occlusion and partially invariant to illumination and the noise present in the image [7]. There are three octaves which will be applied to the image. After each octave, the Gaussian image gets down sampled by factor 2 and then process continues for next two octaves. On completion of this process, we get Difference of Gaussian (DoG) images. For obtaining the maxima and minima of DoG, each selected sample point is compared with its 8 neighbourhood points in current image and 18 neighbourhood points which are present in previous and next image. If the selected sample point is maximum or minimum, than these 27 sample points then it gets selected as a key point which will be used in recognition process. It is highly stable and distinctive in nature than the other points in the neighbourhood. This key point feature localization is used to determine the co-ordinates and the scale present in the image. Some of these key point features have the problem of low contrast and are redundant in nature and thus they get rejected for getting extreme stable points. The Taylor expansion of the scale space function series is used to check these extreme stable points and to reject the poor feature key points.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 y}{\partial x^2} x \qquad (1)$$

where D and its derivatives are evaluated at the sample point. And x= (x, y, σ)$^T$ is offset from this point. The location of the extremum $\hat{X}$ is determined by taking the derivative of this function w.r.t. x and setting it to zero, which gives $\hat{X}$ as follows:

$$\hat{X} = -\left(\frac{\partial^2 y}{\partial x^2}\right)^{-1} \left(\frac{\partial D}{\partial x}\right)$$

(2)

The function value at the extremum helps to reject unstable extrema which has low contrast and it is given by:

$$D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\hat{X}$$

(3)

Here all the pixels get discarded if $|D(\hat{X})|$ < 0.03. But it is not enough to reject the key points with low contrast. The DoG function has a strong response along edges present in the image even though, location along the edge is properly detected, the peaks which are poorly defined on the principal curvature in the DoG function are computed using Hessian matrix. It will have a large value across edge, but small value in the perpendicular direction. In order to eliminate poorly defined key points along the edge, the ratio of principal curvature helps to detect them.

Orientation assignment determines the key point direction to achieve the feature rotation invariance. The direction of the key point can be found by assigning a consistent orientation to each key point based on local image properties. E.g. L(x, y), the modulus value m(x, y) and the orientation Q(x, y) computed using pixel difference is given by:

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

$$\theta(x,y) = \tan^{-1}\left(\frac{(L(x+1,y) - L(x,y-1))}{(L(x+1,y) - L(x-1,y))}\right)$$

(4)

After getting m(x, y) and Q(x, y) the orientation histogram is formed which is later divided in to 36 bins over 0 to 360 degree range of the orientation. Then algorithm corrects and analyses the numerical data of the m(x, y) located in each bin and construct the gradient. Then each key point sample present in the histogram is weighted by Gaussian weighted circular window with an σ which is 1.5 times of the scale of that particular key point. The peaks present in the orientation histogram correspond to the dominant directions of local gradients. Any local peak which is found above the range of 80% of the highest peak in orientation histogram is used to create key point with that orientation. Therefore, many multiple key point can be created at the same location and scale but having different orientation. By finding the gradient orientation and its magnitude, in a region which is around key point location, key point descriptor is produced for each of them. In this way, we get number of key point descriptors of an image. Key point descriptor is in the form of 16 x 16 arrays which contain key points. Each region is divided in to 4 sub-regions of size 4 x 4. Compute the histogram with 8-directional gradient angle for each sub-region, to minimize the effect of illumination normalized histogram. The feature descriptor which is going to be produced length is 4 x 4 x 8 =128 parts.

## IV. Proposed model

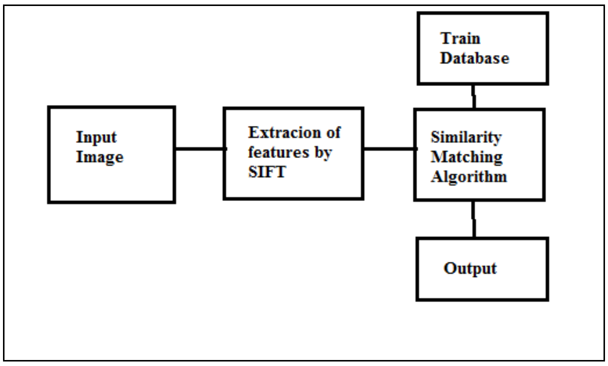



**Fig. 1 Block diagram of proposed system**

In this proposed model there are mainly two sections, feature extraction and recognition. In first section i.e. key point extraction, all the key points are extracted from the images gets stored in the trained database of the proposed system. There are number of steps involved in key point extraction. First step is to create DoG images of the input image, then key point localization and its scale. Only DoG is the efficient function which is used to determine the key point location. DoG function is determined by two adjacent images difference which has different scales. These images which are used to find DoG function are initially multiplying σ to get a smooth and noise free image. σ keeps on incrementing for next 7 images excluding input images. In order to locate only stable and highly distinctive key point, each key point is compared with its 9 key points of the adjacent images which gives the maxima and minima. There will be large key points which get localized but only 15% of them get selected because of they are stable and particularly not affected by noise. In order to detect stable points, set the threshold. Key point features localize along the edges present in the images, are typically unstable to noise, hence we have to reject them. Peaks which are poorly defined can be found at principal curvature. We have to set additional threshold on the principal curvature. After that we detect the stable key points which we use for matching purpose.

## V. Similarity Matching Algorithm

Similarity matching algorithm [9] is computationally complex but it has maximum accuracy while matching one image with the whole database.

$$R(i,j) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}[S(m,n) \times T(m,n)]}{\sum_{m=1}^{M}\sum_{n=1}^{N}[S(m,n)]^2}$$

(5)

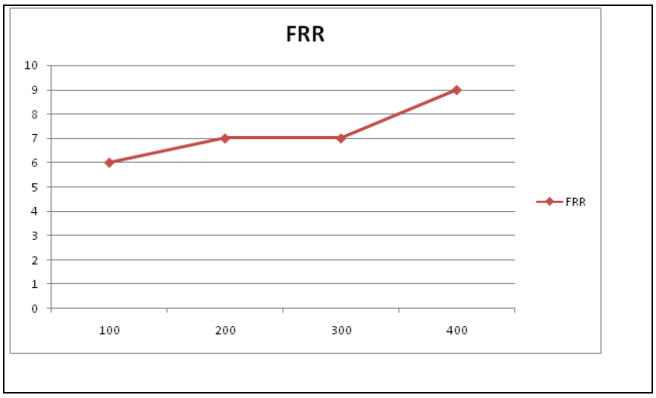
In our proposed system, we have stored all the features of the training images in training database. When input image will be given to the system, its features get extracted using SIFT method, and gets stored in T(m, n). Training image features are stored in S(m, n). When we start computing similarity between training images features with input image features, we calculate R(i, j). As R(i, j) increases towards 1, the probability of getting input image matched with training database also increases. We have set the tolerance for R(i, j). If for any image, R(i, j) becomes greater than 0.90 then that image will be best match with database.

## VI. Experiments and results

This algorithm was tested on the AT&T database which consists of 400 images of 40 different individuals. The database consists of images of these 40 individuals with different pose and expressions. Each image is stored and digitized as 92 x 112 pixel array. The files are in png format. For experimental purpose, we partition 400 images into 2 datasets. First one is trained database and second one is testing database. In training database, first seven images of each individual get stored and thus there are total 280 images and rest 120 images get stored in the test database. While performing the experiment, highly distinctive stable key point descriptors get extracted for all the images in training database and gets stored in training database. As shown in fig. 1 the key points get localized on the input image. Then after applying threshold on principal curvature and selecting the local peaks, which are in the range above 80% of the orientation histogram, only stable key points are allowed and from them we create key point descriptors for matching purpose. As shown in fig. 2, for final recognition step, the key point descriptors of the training database are matched with the test image key point descriptors. This matching process is done by similarity matching algorithm and the accuracy for recognition up to 97% is achieved. The greater the R(i, j) better the results of the algorithm. For matching purpose we have selected the maximum R(i, j) value and the image for which we get maximum value of R(i, j), the image is selected as match for the input image.

## VII. Performance Analysis

Performance analysis of SIFT based face recognition system can be easily studied from the figures given below.






ISSN: 2278 – 1323 

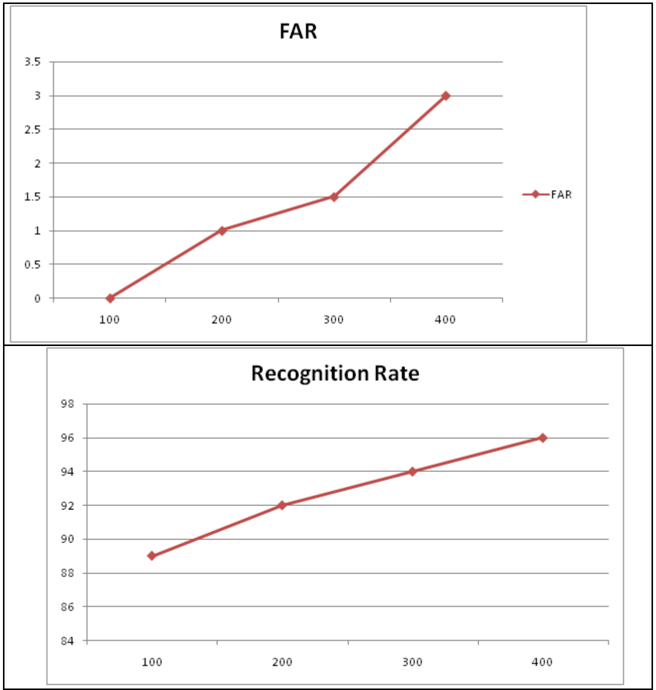



**Fig. 2 (a) FRR (b) FAR (c) Recognition rate**

Above figure shows the recognition rate of the proposed system which is more than 95%. It clearly describes the efficiency of our proposed system.

False acceptance rate (FAR) and false rejection rate (FRR) also been calculated for ORL database. FAR measures falsely accepted users who are not legitimate and FRR measures the true rejection of the user who is the legitimate. From the figure it is clear that as the number of image increases FAR and FRR still remain very low.

Fig. 3 (a) represents the original image. Fig. 3 (b) represents the original image with key points mapped on it and fig. 3 (c) shows the key points allowed for matching purpose.

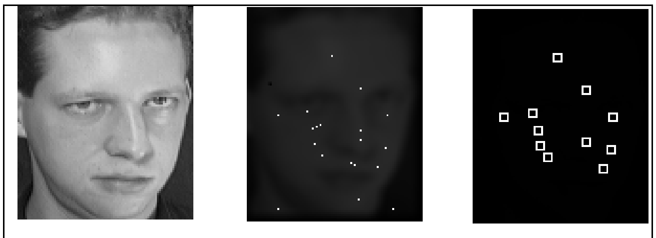


**Fig. 3 (a) Original image (b) Key points localized (c) Key points allowed**

## VIII. Conclusion

Face recognition using SIFT with similarity matching algorithm is implemented successfully in this paper. From the results it is clear that the proposed algorithm gives very high, efficient results under varying illumination conditions, scale, pose and expression. But there is one limitation of this method. In this paper we are using similarity matching algorithm for matching purpose. It is computationally very complex as it is matching each image in the training database pixel by pixel with the input image. So it increases the complexity of the proposed algorithm. Hence in future, there is a scope to reduce its computational complexity

## REFERENCES

[1] Pope, Arthur R. and David G. Lowe, "Learning probabilis-tic appearance models for object recognition," in Early Vi-sual Learning,eds. Shree Nayar and Tomaso Poggio (Oxford University Press, 1996), pp. 67–97.

[2] Beis, Jeff, and David G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces,"Conference on Computer Vision and Pattern Recog-nition,Puerto Rico (1997), pp. 1000–1006

[3] Ito, Minami, Hiroshi Tamura, Ichiro Fujita, and Keiji Tanaka, "Size and position invariance of neuronal responses in monkey inferotemporal cortex,"Journal of Neurophysiol-ogy,73,1 (1995), pp. 218–226.

[4] Cong Geng and Xudong Jiang, "Face Recognition Using Sift Features", ICIP,pp.3313-3316,2009.

[5] S. M. Zakariya, r. Ali and m. A. Lone, "automatic face recognition Using multi-algorithmic approaches", s. Aluru et al. (eds.): ic3 2011, ccis 168, pp. 501–512, 2011.

[6] Sushma Niket Borade and Dr. R.P.Adgaonkar, "Comparrative Analysis of PCA and LDA", ICBEIA,pp,203-206,2011

[7] Wang Yunyi, Huang Chunqing and Qiu Xiaobin, "Multiple Facial Instance for Face Recognition based on SIFT Features", IEEE conference on Mechatronics and Automation,pp.2442-2446,2009.

[8] D. G. Lowe, "distinctive image features from scale-invariant Keypoints", international journal of computer vision, vol. 2, no. 60,Pp. 91-110, 2004.

[9] S. Zhu and K. K. Ma,"A new diamond algorithm for fast block-matching motion estimation," IEEE Trans. Image Processing, vol. 9, pp. 287–290, Feb. 2000.

**Ajitkumar Deshmukh**
M.Tech in Signal Processing, Department of E&TC Engineering, COEP, Pune, Maharashtra. IET member.

**Pritam Sirpotdar**
M.Tech in Signal Processing, Department of E&TC Engineering, COEP, Pune, Maharashtra. IET member.

.