Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2016; 4(1):70-82 **ISSN 2347-9523 (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Research Article

Improved Fuzzy Support Vector Machine for Face Recognition

Navin Prakash[1*], Dr. Yashpal Singh[2]
[1]IFTM University, Moradabad, Uttar Pradesh, India
[2]BIET, Jhansi, Uttar Pradesh, India

***Corresponding author**
Navin Prakash
Email: naveenshran@gmail.com

Abstract: Face recognition is the most emerging research area of the pattern recognition since the early 1990s. This paper presents a new classifier called modified fuzzy support vector machine (MFSVM) by modification in membership function of Fuzzy Support Vector Machine using Combination of Distance Feature, Correlation. We use based on discrete wavelet transform (DWT), discrete Cosine transform (DCT) as a combined feature extraction method and modified fuzzy support vector machine (MFSVM) for face recognition. First the face image is decomposed by 2-D wavelet, then the 2-dimentional DCT is applied to the low-frequency image. Then, using the DCT coefficient, face image can be recognized using MFSVM classifier. The experiment is carried out on the ORL database the result is encouraging, which achieves high accuracy.
Keywords: Face Recognition, discrete wavelet transform (DWT), discrete cosine transform (DCT), fuzzy support vector machine (FSVM), Modified fuzzy support vector machine (MFSVM), feature extraction.

INTRODUCTION

Face recognition has been started since the 1970s .it has wide potential applications such as biometrics systems, surveillance, security control, identity Authentication and human-computer communication etc. Till now, many kinds of face recognition methods exist but broadly face recognition algorithms are divided into three major categories [1]:

Holistic methods

In this approach the whole face area is taken as input data into face grasping system.

Feature-based methods

Feature-based approaches used the local facial features for recognition, such as eyes, nose and mouth.

Hybrid methods

These approaches used both feature-based and holistic features for face recognition. These methods have better performance than individual holistic or feature based method.

LITERATURE REVIEW

A face recognition procedure could be separated into two stages: feature extraction and classification. The feature extraction of a picture is one of the basic errands in picture recognition. Feature extraction algorithms fall into two categories: geometrical features extraction and, Photometric methodologies or statistical features extraction [2].The geometrical methodology, speak to the face as far as structural measurement and distinctive facial features that incorporate distances and angles Between the most trademark face segments, for example, eyes, nose, mouth. Statistical features extraction is normally determined by algebraic strategies, for example, principal component analysis (PCA), what's more, independent component analysis (ICA) [3]. These strategies discover a mapping between the original feature spaces to a lower dimensional feature space. The lack of PCA is that it treats inner-class and out-class similarly [4-6] and in this way it is sensitive to light and changes of expressions. Transformation-based feature extraction methods, for example, the DCT and DWT used to concentrate highlights with extract features with very low computational cost [7].DCT used as a part of image compressing which is moreover used to extract features [8-9].Wavelet analysis has both decent qualities in time domain and frequency domain for unsteady signals with high precision.

Here we are using a combined feature extraction method based on DWT and DCT and proposed a modified fuzzy support vector machine (MFSVM) for classification of faces.

Feature extraction using discrete wavelet transform and discrete cosine transform

Discrete wavelet transform on face image

Different frequency bands assume distinctive part in face recognition. Little variety of face expression will basically influence high-frequency components Therefore. Discrete Wavelet transform isolates a picture into four different frequency bands in every level of decomposition. Discrete Wavelet transform for face picture have a few focal points [12]:

1. By breaking down face picture utilizing wavelet transform, four different frequency bands, namely low-low frequency band LL, low-high frequency band LH, high-low frequency band HL and high-high frequency band HH are accomplished and for further breaking down the low-frequency picture for accomplishing the global and details of face picture.

2. If n-level of wavelet decomposition is applied to the face picture, the resolution of the low-frequency subband is extraordinarily not exactly the resolution of the original face picture and it lessens the space dimension of the original data.Fig.1 demonstrates that the decomposition process by applying the 2D wavelet transform on a face picture.

Here we break down one face picture into two level utilizing Daubechies wavelet. The original picture can be decomposed into four subband pictures in every level of decomposition.

(a) (b) (c)

Fig-1: 2D- wavelet decomposition of image in ORL database (a) Original face and (b) level 1 db2 wavelet decomposition into Subband, upper-left comer is the sub band LL (c) level 2 db2 wavelet decomposition

2D DWT can be got as:

(1)Use 1-D DWT row-wise (to generate L and H sub bands for each row) at that point column-wise in the first level of decomposition as given in figure 2.we will develop four sub-bands LL1, LH1, HL1 and HH1 are found. Then repeating the process in the LL1 subband, it will generate LL2, LH2, HL2 and HH2 as given in Fig. 2.

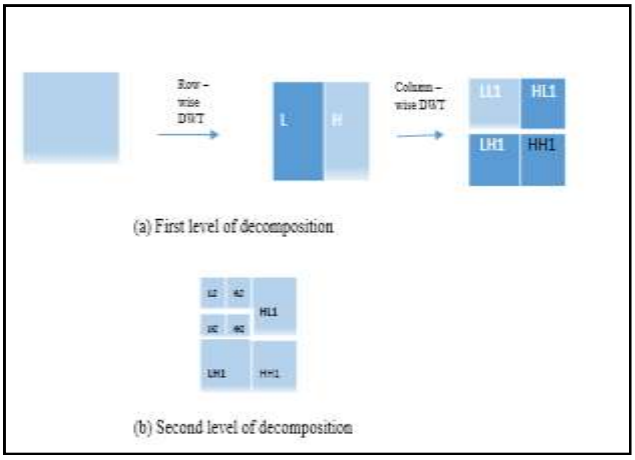


Fig-2: Wavelet decomposition process

Wavelet theory gives, high frequency is more sensitive to facial variation so HH is most sensitive while LL features are less sensitive to the facial expressions. So all LL subband coefficients are selected to recognition .The DWT is strictly related to multi-resolution analysis and sub-band decomposition. The 1D-DWT recursively decomposes the input signal, So(n), in approximation and detail at the next lowest resolution stages. If Si(n) is the approximation and Wi(n)is the detail, of the signal at level i. The approximation of the signal at level i+ 1 is calculated using equation [10]:

$$S_{i+1}(n) = \sum_{K=0}^{L-1} g(k)\, s_i(2n-k) \qquad (1)$$

And the detail of the signal at level i+ 1 is calculated as:

$$w_{i+1}(n) = \sum_{K=0}^{L-1} h(k)\, s_i(2n-k) \qquad (2)$$

Where g (k) is the low-pass and h (k) is high-pass filter coefficients, and L is the size of filters. This method for calculating the DWT is normally denoted to as the Mallat's algorithm [9].The 2 D-DWT functions on 2-D signals, such as images. Whereas 1-D filters are applied to calculate the 1-D-DWT, the 2-DDWTapplied 2-D- filters in its calculation. A 2-D filter f(nl, n2) is separable if it can be described as f(n1 , n2) =f1(n1) f2(n2) , wheref1(nl) and f2(n2) are 1-D filters. The separable 2 D-DWT decomposes an approximation image Si (nl, n2) into an approximation image and three detailed images giving to:

$$S_{i+1}(n_1,n_2) = \sum_{K1=0}^{L-1}\sum_{K2=0}^{L-1} g(k_1)g(k_2)S_i(2n_1-k_1,2n_2-k_2) \qquad (3)$$

$$W_{i+1}^{1}(n_1,n_2) = \sum_{K1=0}^{L-1}\sum_{K2=0}^{L-1} g(k_1)h(k_2)S_i(2n_1-k_1,2n_2-k_2) \qquad (4)$$

$$W_{i+1}^{2}(n_1,n_2) = \sum_{K1=0}^{L-1}\sum_{K2=0}^{L-1} h(k_1)g(k_2)S_i(2n_1-k_1,2n_2-k_2) \qquad (5)$$

$$W_{i+1}^{3}(n_1,n_2) = \sum_{K1=0}^{L-1}\sum_{K2=0}^{L-1} h(k_1)h(k_2)S_i(2n_1-k_1,2n_2-k_2) \qquad (6)$$

The signal S i+1 (nl, n2) is called as low-low (LL) subbands which is an approximation of Si(nl, n2) at a lower resolution. This approximation is calculated from S, (nl, n2) by low-pass filtering and decimating by 2 along its rows and columns. The signals W_{i+1}^{1}(n1,n2),W_{i+1}^{2} (n1,n2)andW_{i+1}^{3}(n1,n2) are correspondingly named as Low-high(LH),high-Low(HL) and high-high(HH) sub bands which hold the detail of Si(n1,n2).The one-level 2D-DWT calculation is shown in Fig. 3.

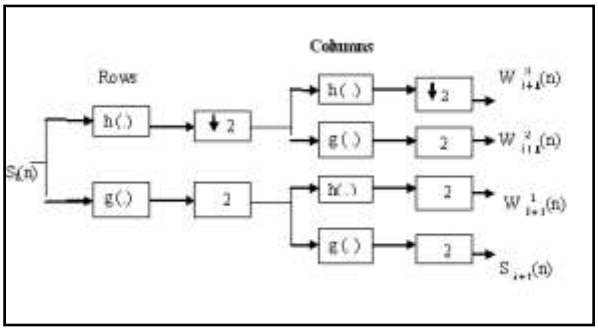


Fig-3: Block Diagram of the analysis filter bank used to compute the 2D-DWT

The sub band LL is the low frequency sub band of the original Picture and it contributes to the global sketch ofa face; While LH,HL,HH are the high frequency sub band and they hold broad information of a face. The sub band LL is the greatest stable sub band for features of a face image.

Discrete cosine transform

The DCT translates high –dimensional face image into low dimensional face picture which contains important features. The common equation for the DCT of an M x N image f(x, y) is defined by the equation (7) given below [11]:

$$F(u,v) = a(u)a(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} \cos\left[\frac{\pi u}{2M}(2x+1)\right]\cos\left[\frac{\pi v}{2N}(2y+1)\right]f(x,y) \qquad (7)$$

Where u=0, 1, 2… M-1 and
V=0, 1, 2… N-1. a (u), a(v) are defined as

$$a(u) = \begin{cases} \frac{1}{\sqrt{M}} u = 0 \\ \frac{2}{\sqrt{M}} u \neq 0 \end{cases} (8)$$

$$a(v) = \begin{cases} \frac{1}{\sqrt{N}} v = 0 \\ \frac{2}{\sqrt{N}} v \neq 0 \end{cases} \quad (9)$$

The DCT coefficient for an image with large magnitude is generally situated in the upper-left comer in DCT matrix. We can use correct DCT coefficient from the upper-left comer and then convert it to a one-dimensional vector which will be the input of a classifier.

SVM AND Fuzzy SVM

Support Vector Machine

Support Vector Machine is a classifier that often determines optimal hyperplane in input space by minimizing the average error in training data and maximizes the margin between classes. The maximum margin hyperplane is the one that gives the greatest separation between the classes. And nearest data in every class is called as support vectors. As notice in figure 4.

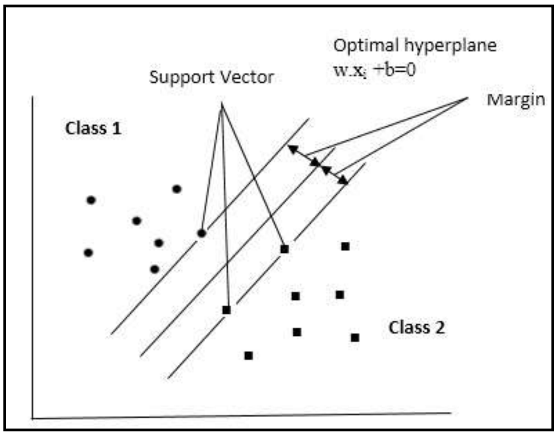


Fig-4: Optimal hyperplane

Suppose there are n datasets {x1, ….., xn}, where$x_i \in R^n$ belongs to one of the two classes as yi$\in$ {+1, -1}, after that each sample point must to satisfied the given conditions as:

$$w.x_i+b \geq 1 \text{ for } y_i = +1$$
$$w.x_i+b \leq 1 \text{ for } y_i = -1 \quad (10)$$

Where w is a normal to the hyperplane and b called bias of the hyperplane. Then, the margin of separating hyperplane defined as:

$$\frac{1-b-(-1-b)}{w} = \frac{2}{|w|} \quad (11)$$

Then maximizing $\frac{1}{|w|^2}$ is equivalent by minimizing,$|w|^2$then the optimal hyperplane can obtain by:

$$\min \frac{1}{2} \|w\|^2 \quad (12)$$
$$\text{s.t} \quad y_i(w.x_i+b)-1 \geq 0$$

For non-separable case, we present nonnegative slack variables ξ such that the given conditions are satisfied as :

$$\min \frac{1}{2} \|w\|^2 + C(\sum_{i=1}^{n} \xi_i) \quad (13)$$
$$\text{s.t.} \quad y_i(w.x_i+b)-1 \geq 1 - \xi \quad , \xi \geq 0$$

Where ξi =0, if xi is correctly classified. C is a positive constant which defines the penalty on empirical error.

The above problem is a quadratic problem that is hard to solve and needs a heavy computation time. Then, a set of Lagrange multipliers (α) is present and the problem is changed into

$$\min_{w,b} L_P(w,b,\alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \alpha_i \, y_i(x_i + b) + \sum_{i=1}^{n} \alpha_i \quad (14)$$

These points with αi > 0 is named as support vector. If 0< αi < C, these points are lie on one of the margin. Whereas if αi = C, these points are supposed to be as misclassified data.

Minimizing the Lp with respect to b ($\frac{\partial}{\partial b} L_P(w, b, \alpha)$ =0)and w ($\frac{\partial}{\partial w} L_P(w, b, \alpha) = 0$) , then we got:

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \ ,$$

$$W=\sum_{i=1}^{n} \alpha_i y_i x_i \qquad (15)$$

The W vector is possibly infinite, but the value of αi is finite. Henceforth, by replacing the equation above we will be found the dual problem

$$L_P(\alpha) \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \qquad (16)$$

So to find the optimal hyperplane defined as:

$$maxL_P(\alpha) \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \qquad (17)$$

$$\text{s.t. } \sum_{i=1}^{n} \alpha_i y_i = 0, 0 \le \alpha_i \le C$$

Non-separable case additionally can be comprehended by kernel idea. This original input space is mapped into a high-dimensional dot product space named as feature space. The mapping illustration shown as below in figure 5.

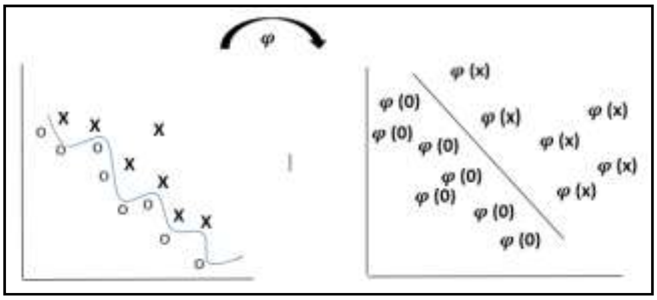


Fig-5: Mapping by kernel trick

Following Kernel functions can be used in pattern recognition as:

Linear: k (x_i, x) =x_i^T
Polynomial: k $(x_i, x) = (\gamma x_i^T x + r)^P, \gamma > 0$
Radial Basis Function (RBF): k (x_i, x) =exp $(-\gamma |x_i - x|^2), \gamma > 0$
Tangent hyperbolic (sigmoid):k (x_i, x) =tanh$(\gamma x_i^T x + r)$

Fuzzy Support Vector Machine

Fuzzy SVM was first given by Shigeo Abe in 2001 to focus the unclassifiable areas issue. By then, in 2002, Chun-Fu Lin utilized Fuzzy SVM to diminish the impact of outliers in model generation. In normal SVM, the sample points are managed comparatively, including outlier and every point is thought to be had a place with one class. In any case, in some genuine issue, some input points may not accurately delegate to one of these classes and distinctive points have a different responsibility to the choice surface. For this issue, fuzzy membership is presented by relegating a membership value μ_i to every training point x_i.The higher value of μ_i makes the comparing point xi less vital in the training, and vice-versa.

If n dataset with fuzzy membership {$(x_1, y_1, \mu_i)\ldots(x_n, y_n, \mu_n)$ }, where $x_i \in R^n$ belongs to one of the two classes as yi∈ {+1, -1}these classes have fuzzy membership $t \le \mu_i \le 1$ with adequate small t > 0. At this point, the separating hyperplane defined as:

$$\min \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n} \mu_i \xi_i\right) \qquad (18)$$

s.t. y_i(w.x_i+b)-1$\geq 1-\xi_i$,$\xi_i \geq 0$
so,the problem can be transformed into Lagrange function as:

$$maxL_P \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \qquad (19)$$

$$\text{s.t. } \sum_{i=1}^{n} \alpha_i y_i = 0{,}0 \leq \alpha_i \leq \mu_i C$$

In standard SVM, there is simply free parameter C which is used to control the exchange off between maximizing margin and minimizing the measure of misclassification. A greater C makes the training SVM less misclassification and the margin will be smaller. The decrease of C makes SVM ignore all the all the more training points and get more extensive margin. The more prominent the estimation of C, then a couple of blunders and the margin will be littler. The diminished the estimation of C will be the reason for blunders and the margin considerably more prominent.

In fuzzy SVM, the number of free parameters is equivalent to the number of training points because each training points is given a membership value to control the trade-off. The membership value is usually calculated based on the main property of dataset. As we can use fuzzy membership as function of the distance between the point and its class center.

Proposed membership function for Fuzzy SVM

Selecting suitable fuzzy membership for a problem is very significant in FSVM [15].As we know that the different membership functions have different effect to the algorithm. This paper introduce new membership function by combining of distance feature, covariance and clustering algorithm. The distance concept identify the overlapping between classes, while the covariance is used to represent the similarity between points. Then, we will calculate the fuzzy membership for each feature.

Distance

Most researchers define fuzzy membership as a function of distance between each sample point to its class center [14]. So, fuzzy membership can be defined as a function of the distance between the point and class center, then the fuzzy membership defined as:

In most paper, fuzzy membership often define as a function of distance between each sample point to its class center [14]. This is one of important basis for measuring contributions which sample point have.

In this paper, we use this concept to identify the overlapping between classes [16]. Therefore, we are not just only computed each point to its class center, but to all class center. This concept is shown in figure 6.

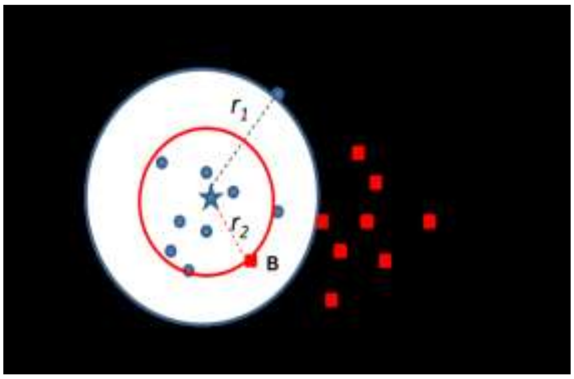


Fig-6: The distance concept

Figure 3, shows the observations to the class+. Point A is the farthest point of class+, while point B is the closest point from another class (class -) to class+ center. Thus, the distance between the class+ center (x+) and point A is called as maximum radius (r1), whereas the distance to point B is called as minimum radius (r2). If r1> r2 means there is overlapping between classed. So, the fuzzy membership function based on distance (μ_1(x)) is defined as

$$\begin{cases} \mathbf{1}, & \boldsymbol{if}\ \|\boldsymbol{x}_+ - \boldsymbol{x}_i\| < \boldsymbol{r}_2 \\ \mathbf{1} - \frac{\|x_+ - x_i\|}{r_1}, & \boldsymbol{if}\ \boldsymbol{r}_2 \le \|\boldsymbol{x}_+ - \boldsymbol{x}_i\| \le \boldsymbol{r}_1 \end{cases} \quad (19)$$

In this paper, the class centers are calculated by Hodges-Lehmann (HL) method. This method is a robust statistical method to find the class center based on median of the average data pairs, as expressed as follow:

HL = median $x_i - x_j/2$, $1 \le i \le j \le n$ (20)

where n is the number of data. HL estimator is a more reliable than using the mean value. The mean is very sensitive to outliers because this function calculates the average value of an entire class. For example, if there is a point which have a wide range value compared with other data, then the average value of the class can be much different. Meanwhile, HL estimator determines the location of the center of classes based on ranges of data values in a class. The outermost data contribution can be Calculate using the cluster whose center is closest to that point.

Similarity

The similarity is the second variable used to find out the fuzzy membership. Covariance indicates how two variable are related and correlation is also able to measure the degree to which the variables tend to move together. One of the methods to amount the similarity of two data is using the correlation using covariance as:

Cov $(\boldsymbol{x_i},\boldsymbol{x_j})=$

$$\frac{\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x_j} - \bar{\boldsymbol{x}}_j)}{\boldsymbol{n} - \mathbf{1}}$$

Sim $(\boldsymbol{x_i},\boldsymbol{x_j})$ =corr $(\boldsymbol{x_i},\boldsymbol{x_j})$

$$=\frac{\text{cov}\ (x_i, x_j)}{s_x . s_y}$$

Where

s_x = sample standard deviation of the random variable x

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}_i)^2}{n-1}}$$

s_y = sample standard deviation of the random variable y

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{x_j} - \bar{\boldsymbol{x}}_j)^2}{n-1}}$$

The data with low similarity will be considered as outlier. In this paper, we use correlation to find the similarity between each point to its class center. The fuzzy membership function based on similarity $\mu_2(x)$ is defined as

$$\mu_2(x) = \begin{cases} 1 & if\ abs(corr(x_i, x_+)) \ge 0.9 \\ abs(corr(x_i, x_+)) & if\ abs(corr(x_i, x_+)) < 0.9 \end{cases} \quad (21)$$

If correlation coefficient more than 0.9 means having a very strong relationship. Therefore, we define the degree membership for such point is 1. Otherwise, correlation will be close to 0 if there is no relationship exists.

Clustering

We use fuzzy clustering methods to decide clusters. These clusters comprise both normal and outlier data points. Set 1 Fuzzy memberships of these data points and fuzzy memberships of other data points are calculated by using the closest cluster.

Briefly, the algorithm of membership functions generation can be written as the following stages:

Algorithm-

Step1

Check whether there is missing value in each attribute. If missing value found, then substitute by mean value of its attributes.

Step 2
For each class, find the class center (c) and the outermost point.

Step 3
Find the membership degree based on distance ($\mu_1(x)$).
For each class *i*, find the minimum and maximum radius. Minimum radius is the distance from the class center *c* to the outermost point from the other class (class *j*), whereas the maximum radius is the distance from the class center to the outermost point of the class *i*. Then, calculate the distance of each point in class *i* to its class center. Apply the membership degree according to equation 14.

Step 4
Find the membership degree based on Pearson correlation ($\mu_2(x)$).
For each point, calculate its correlation to class center and apply the membership degree according to equation 16.

Step5
To find out membership degree based on clusters ($\mu_3(x)$). Perform clustering on the training data set.
a) Select a clustering algorithm
b) Perform clustering on the training data set
c) Determine a subset containing clusters that contain both normal and abnormal data. Denote this subset as ALLCLUS.
d) For each data point x $\in$ ALLCLUS, set its fuzzy membership to 1
e) For each data point x$\notin$ALLCLUS, do the following
i.Find out the cluster whose center is closest to x
ii.Calculate fuzzy membership of x with this cluster

Step6: Find the membership degree based on different observations is combined by taking mean and obtain membership degreeμ(x)**.**

$$\mu(x) = \text{mean}\,[\mu_1(x), \mu_2(x), \mu_3(x)]$$

Table-1: Diffrent membership degree

$\mu_1(x)$	$\mu_2(x)$	$\mu_3(x)$	$\mu(x)$
0	0	0	0
0	0	1	0.3333
0	1	0	0.3333
0	1	1	0.6666
1	0	0	0.3333
1	0	1	0.6666
1	1	0	0.6666
1	1	1	1

$$\mu(x) = \begin{cases} 1 & , if\, \mu(x) \geq 0.6666 \\ 0 & , Otherwise \end{cases}$$

Face Recognition

Face recognition is generally divided into training process and recognition process. Figure 7 shows the diagram of face recognition based on DWT/DCT and MFSVM

A. Training Process

Step 1: The 2-dimentional discrete wavelet transform is applied to the input image. The low frequency sub band acts as the approximation of the input image.

Step 2: The 2-dimentional DCT is applied to the low frequency approximation image obtained from step l.

Step 3: Build FSVM classifier based on the DCT coefficients got from step

B. Recognition process

The same process (training process's step 1 and step 2) is carried out on the testing image and then recognize face image using MFSVM classifier.

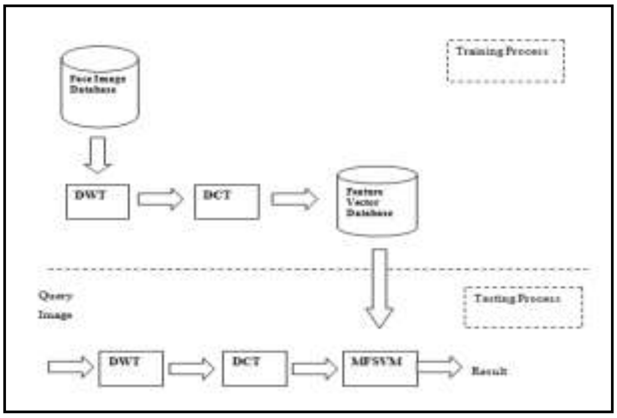


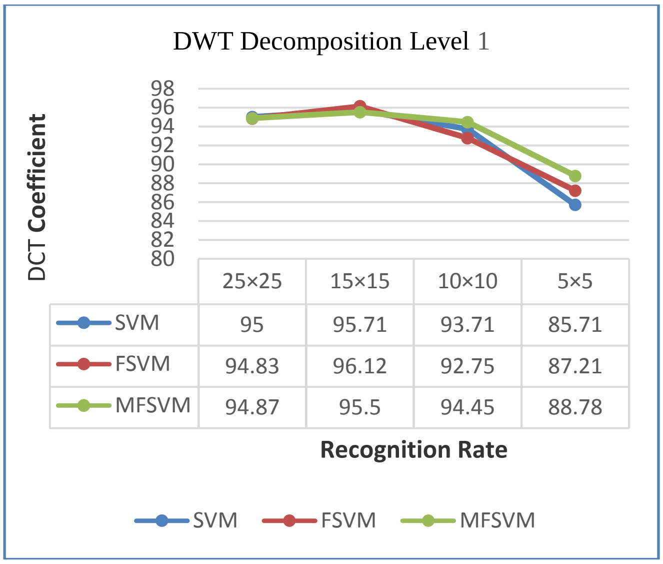
Fig-7: Proposed face recognition

Experimental Results and Analysis

The ORL face database composed of images of size 112 x 92 resolution. The training set was set up by a random selection of four images and a testing set of the other six images. 400 tests were taken in our scheme with 18 false recognition with four samples. We applied Db2 as the mother wavelet. We performed 1-level, 2-level, and 3 -level decomposition respectively and the dimensions of the sub band LLl, LL2, LL3 is 57×47, 30×25, and 16×14. The 2-dimentional DCT is applied to the low frequency image obtained from DWT and only a subset of the DCT coefficients corresponding to the upper left comer of the DCT array is retained. Subset sizes of 25×25, 15×15, 10×10 and 5×5 of the original DCT array are used in this experiment as input to the different classifier as given in table 2.

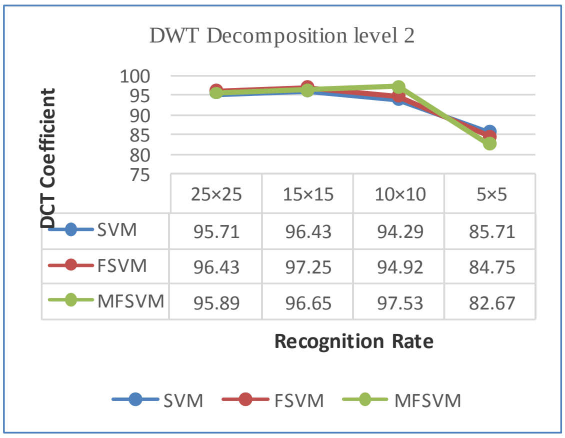
Table-2: Experimental Result with 4 samples

DWT Decomposition Level	DCT Coefficient	Recognition Rate		
		SVM	FSVM	MFSVM
1	25 ×25	95.00	94.83	94.87
1	15 × 15	95.71	96.12	95.50
1	10 × 10	93.57	92.75	94.45
1	5 ×5	85.71	87.21	88.78
2	25 × 25	95.71	96.43	95.89
2	15 × 15	96.43	97.25	96.65
2	10× 10	94.29	94.92	97.53
2	5 ×5	85.71	84.75	82.67
3	10 ×10	93	95.29	95.81
3	5 ×5	87.86	86.76	91.43



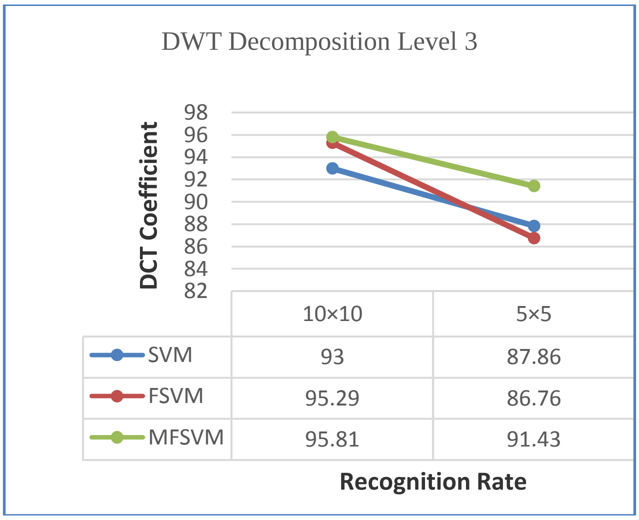
	25×25	15×15	10×10	5×5
SVM	95	95.71	93.71	85.71
FSVM	94.83	96.12	92.75	87.21
MFSVM	94.87	95.5	94.45	88.78

Fig-8: Recognition Rate at Decomposition Level 1



	25×25	15×15	10×10	5×5
SVM	95.71	96.43	94.29	85.71
FSVM	96.43	97.25	94.92	84.75
MFSVM	95.89	96.65	97.53	82.67

Fig-9: Recognition Rate at Decomposition Level 2



	10×10	5×5
SVM	93	87.86
FSVM	95.29	86.76
MFSVM	95.81	91.43

Fig-10: Recognition Rate at Decomposition Level 3

We can find out that experimental result is shown in table 2, and as shown in figure 9, the recognition rate reaches 96.43% when the decomposition level is 2 and DCT coefficient is 15×15. At the same time, an experiment carried out using the FSVM and MFSVM on the ORL face database recognition rate is 97.25% and 96.65% respectively.

And in the same figure shown that using MFSVM at decomposition level is 2 and DCT coefficient is 10× then recognition rate is 97.53.So we could come to the conclusion that MFSVM is superior to traditional SVM and FSVM.

The training set was set up by a random selection of six samples per person from the whole database and the testing set was the remaining images.

As given in Table- 3 shows the recognition results. 400 tests were taken in our scheme with 8 false (6 samples).

Table-3: Experimental Result with 6 samples

DWT Decompostion Level	DCT Coefficient	Recognition Rate		
		SVM	FSVM	MFSVM
1	25×25	93.45	93.72	93.78
1	15×15	93.38	96.50	95.75
1	10×10	94.25	93.12	94.25
1	5×5	86.34	87.45	89.27
2	25×25	95.76	94.84	93.87
2	15×15	94.29	96.35	96.60
2	10×10	95.26	95.32	97.90
2	5×5	83.75	82.19	82.85
3	10×10	93.45	95.57	96.29
3	5×5	88.12	82.49	90.43

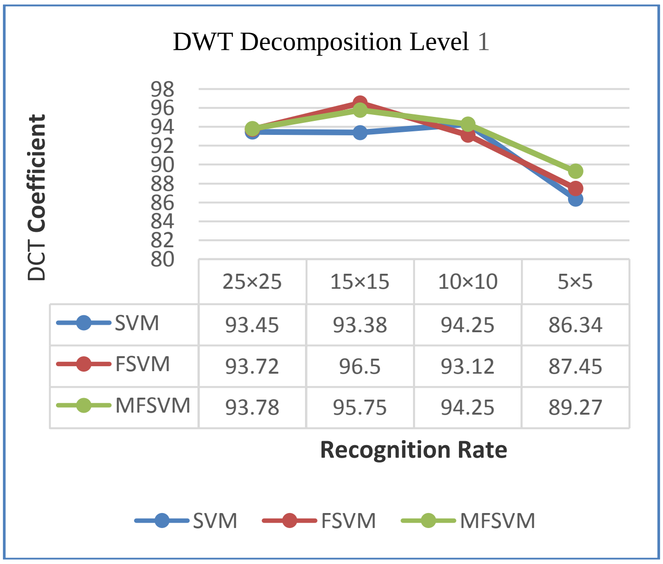


Fig-11: Recognition Rate at Decomposition Level 1

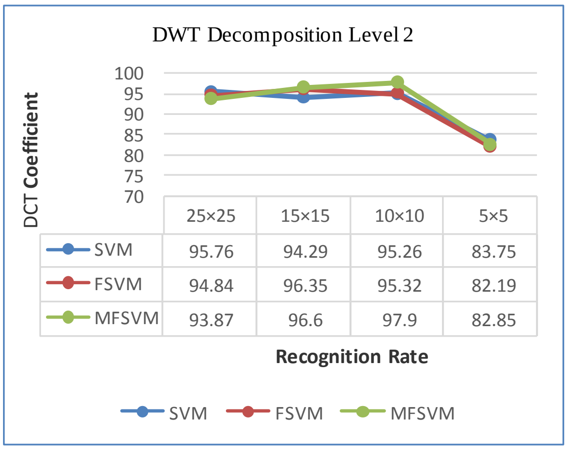


Fig-12: Recognition Rate at Decomposition Level 2

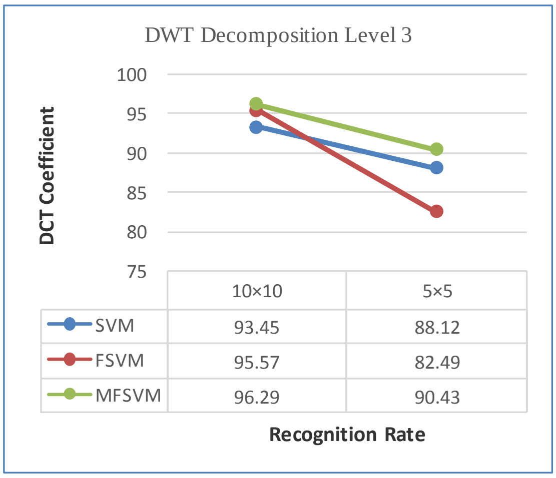


Fig-13: Recognition Rate at Decomposition Level 3

We can find out that experimental result is shown in table 3, and As shown in figure-11, the recognition rate reaches 95.75% when the decomposition level is 1 and DCT coefficient is 15×15. and experiment carried out when decomposition level is 2 and the DCT coefficient is 15×15 FSVM recognition rate is 96.50% and DCT coefficient is 10 ×10 at same decomposition level MFSVM gives recognition rate 97.90% on the ORL face database. And in the same figure shown that using MFSVM at decomposition level is 2 and DCT coefficient is 10× 10 then recognition rate is 97.90.So we could come to the conclusion that MFSVM is superior to traditional SVM and FSVM.

Conclusion and future Direction

In this paper, a new fuzzy membership function proposed for FSVM using the combination of two variables: distance, similarity. It gives better performance on reducing the effects of outliers than some existing methods. This paper presents a DCT and DWT Mixed approach for feature extraction during the face recognition, the Modified Fuzzy Support Vector Machine (MFSVM) is tested for wide facial variations. The experiments that we have conducted on the ORL database show that the MFSVM method achieves excellent performance than FSVM in terms of recognition rates.

Here remains some future work to be done as Combine with the other subband coefficients of wavelet decomposition to improve recognition ability is required to be studied future. Wavelet transform and multisolution analysis can reduce image data greatly and extract main face information which is adequate for recognition. The ultimate goal is to automatically determine an appropriate fuzzy membership function that can reduce the effect of noises and outliers that can help face recognition in different situations. One of the great issues in FSVM that how to reduce the time in solving face recognition.

Acknowledgments

We would like to thank Dr. Yashpal Singh, Associate Professor, BIET-Jhansi, for his valuable guidance. I also thankful to respected Vice-Chancellor, Dr.R.M Dubey sir for their moral support. And I also wish to express our gratitude to my parents, who render their support during the period of research work.

REFERENCES

1. Zhao W, Chellappa R, Phillips PJ, & Rosenfeld A; Face recognition: A literature survey. ACM computing surveys (CSUR), 2003; 35(4): 399-458.
2. Rouhi R, Amiri M, Irannejad B; AReview ON FEATURE EXTRACTION TECHNIQUES IN FACE RECOGNITION. Signal & Image Processing, 2012; 3(6): 1.
3. Yi-qiong X, Bi-Cheng L, Wang B; Face recognition by fast independent component analysis and genetic algorithm. In Computer and Information Technology, 2004. CIT'04. The Fourth International Conference on, IEEE, 2004;194-198.
4. Liu C, Wechsler H; Evolutionary pursuit and its application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000; 22(6): 570-582.
5. Turk M, Pentland AP; Face recognition using eigenfaces. In Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE, 1991;586-591.
6. Du L, Jia Z, Xue L; Human face recognition based on principal component analysis and particle swarm optimization-bp neural network. In Natural Computation, 2007. ICNC 2007. Third International Conference on, IEEE, 2007; 3: 287-291.
7. Samra AS, El Taweel Gad Allah SG, Ibrahim RM; Face recognition using wavelet transform, fast fourier transform and discrete cosine transform. In Circuits and Systems, 2003 IEEE 46th Midwest Symposium on, IEEE, 2003; 272-275.
8. Podilchuk C, Zhang X; Face recognition using DCT-based feature vectors. In Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, 1IEEE, 1996;4: 2144-2147.
9. Yankun Z, Chongqing L; Efficient face recognition method based on DCT and LDA. Systems Engineering and Electronics, Journal of, 2004; 15(2): 211-216.
10. Colom-Palero RJ, Gadea-Girones R, Ballester-Merelo FJ , Martínez-Peiro M; Flexible architecture for the implementation of the two-dimensional discrete wavelet transform (2D-DWT) oriented to FPGA devices. Microprocessors and Microsystems, 2004; 28(9): 509-518.
11. Yin HT, Fu P, Sha XJ; Face recognition based on DCT and LDA. Acta Electronica Sinica, 2009; 37(10): 2211-2214.
12. Liu Y, Chen G, Lu J, Chen W; Face recognition based on independent component analysis and fuzzy support vector machine. In Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on, IEEE, 2006; 2: 9889-9892.
13. Cortes C, Vapnik V; Support-vector networks. Machine learning, 1995; 20(3): 273-297.
14. Lin CF, Wang SD; Fuzzy support vector machines. Neural Networks, IEEE Transactions on, 2002; 13(2): 464-471.
15. Jiang X, Yi Z, Lv JC; Fuzzy SVM with a new fuzzy membership function. Neural Computing & Applications, 2006; 15(3-4): 268-276.