Face Gender Image Classification Using Various Wavelet Transform and Support Vector Machine with various Kernels.

Amjath Fareeth Basha[1], Gul Shaira Banu Jahangeer[2]

[1] College of Computers and Information Technology, Taif University, Taif, KSA

[2] Computers and Information Technology, Taif University, Taif, KSA

Abstract

When we look at a face, we readily identify that person's gender (male/female), expression (happy /unhappy), personality (He / She), age, and charisma. Gender classification such as classifying human face is only challenging for computer, but even hard for human in some cases. In this paper a new novel approach is proposed to recognize gender from the face image. Continuous Wavelet Transforms are used for features selections for each face images of male and female. These selected features will be used to classify the face images of each Gender using Support Vector Machine with Linear Kernel. This Paper use ORL database contain 400 images include both Male and Female Gender .The experimental result shows that the proposed approach (Continuous wavelet Transform and Support Vector Machine) achieves excellent classification accuracy (98%)compared with other Technique like Discrete wavelet Transform and Radon Transform with Support Vector Machine .

Keywords: *Face Gender Classification, Feature Selection, Continuous Wavelet Transform (CWT), Discrete Wavelet Transform, Radon Transform. Support Vector Machine (SVM)*

1. Introduction

Gender classification using facial images has been in the field of research now days and it is quite interesting. Humans are very good in differentiating the gender from facial images. Social Behavior and human interaction is mainly depending upon on the gender of the person with whom he/she they plan to interrelate. Luckily, human being has the unique capacity of classify gender analyzing simply ones face and exclusive character of conveying personality, emotions, age and lot of other vital information. Even if the face of the human is damaged, to find the gender symptoms, we can identify the gender with very high accuracy [1]. More recently automated gender classification from facial images has gained much interest in computer vision, machine language and Image processing community. The Rapid development progress of this gender classification research area is due to the fast growing field of Internet, electronic commerce, electronic banking systems, more human computer interaction, and demographic research, and security and surveillance applications. It can also bump up other important areas like Image /video indexing, retrieval, passive demographic data collection, vision based human monitoring, human robot interaction , face recognition, face detection, age, traditional determinations are some other important application, where gender classification play a major role.

Automatic gender classification is more challenges now days and it's mainly difficult because of the inherent variability of human faces due to different image formation process in terms of image quality, photometry, geometry, occlusion etc [1]. Gender classification has been studied less. Gender classification has been especially interesting for psychologists but automatic gender classification has applications also in other fields, for example, in demographic data collection [2].Automatic gender classification is also a useful preprocessing step for face recognition since it is possible to halve, in case of equal amount of both genders, doing the face classification before the face recognition will make the face recognition process almost twice as fast. In addition, separate face recognizers can be trained for the genders and in this way increase the face recognition accuracy. This has been successfully experimented in facial expression recognition [3]. Erno Mäkinen and Roope Raisamo provide the guidelines for classification with automatic detection and aligned faces [4].

Many types of gender classification methods are available appearance – based method or holistic approach, geometric or feature based approach, hybrid approach [1].in appearance based approach, the whole image or specific regions in a face images is used to generate feature vector .Feature based approach need to localize differential components such as eyes, nose, eyebrows etc. in hybrid approach perceives both local features and whole face.

Many techniques have been taken to classify facial images based on gender. This paper works out on the particular approach using Continues Wavelet Transform (CWT) and Support Vector Machine (SVM) for classifying the gender of the facial images and compared with DWT, RADON along with SVM.

To analyze all the features describing an image and to detect gender of the images, it is important to extract all the available gender information from the image. It can be helpful to analyze the image at different resolution levels. Wavelet transform is an ideal tool to analyze images of different gender. It discriminates among several spatial orientations and decomposes images into different scale orientations, providing a method for in space scale representation. The general principles of wavelet transforms have been described elsewhere [5]. Wavelet functions can be used to select the important features for gender classification. In this paper Continuous wavelet Transform have been applied to gender images with varying success. Many authors have developed computerized methods to classify gender face images. In our Proposed Method along with Continuous Wavelet Transform, We Classify the gender using Support Vector Machine Our technique performs over well in images containing variations in lighting and facial expression, pose angles, aging effects etc. Moreover it is less time consuming process when we compared with discrete wavelet Transform, Radon Transform along with Support vector Machine.

In section 1, we introduce the goals of the paper. Section 2 describes the proposed technique. The Feature Selection using DWT, RADON, CWT is presented in Section 3.Section 4 discusses classification and prediction using SVM. Finally, some experimental results with discussion and conclusions are given in section 5 and 6 respectively.

2. Experimental Setup

In this section, we describe our experimental setup, which explain about facial dataset and the proposed novel technique with raw data, and analyze the informative features in images containing variations in lighting and facial expression, pose angles, aging effects etc and finding out the wavelet coefficient using Continuous wavelet transform in gender facial images . One dimensional wavelet transform performs better in images containing variations in lighting and facial expression, pose angles, aging effects etc. Moreover it is less time consuming process. Classification and Prediction is done with the Support Vector Machine Classifier. It classifies the gender as male and female with various constrains and do better prediction.

2.1 Dataset Description

The paper uses the image dataset called ORL Database. The ORL database totally consists of 400 gray scale images representing male and female gender. This images contains variations in lighting, facial expressions, pose, angles, age effects information. In this work, we collect 400 face images out of which 350 faces are male and rest 50 images are female.

Figure 1. ORL Database of 400 images

2.2 The Proposed Technique

The proposed technique includes application of 1-D Continuous wavelet transform (CWT) on the facial images. The resultant wavelet coefficients are sorted, and take 100 coefficients in file. During the training phase, the SVM is trained with 400 images in the dataset that includes variables pertaining to the corresponding coefficients with label 0 for male and label 1 for female images. Similarly, 200 images is used in testing phase and apply the SVM classification model on the testing coefficients from the CWT of test image set. Finally, SVM prediction rate is calculated in terms of Mean Squared Error (MSE).Thus the classification of facial images is achieved in two main steps. In the first step, features are selected from the gender facial images. In the second step, the selected features with higher wavelet coefficients with label 0 or 1 are classified using SVM classifier.

The steps of the proposed technique are as follows:

Step 1: Read an image one by one.
Step 2: Convert the image into single dimensional array.
Step 3: Apply the 1-D Continuous wavelet transform (CWT).
Step 4: Take the coefficient of all images, which is consider along with the label (0 for male, and 1 for female)

Step 5: Repeat the Steps 1 through Step 4 for all the images.
Step 6: Train the dataset using support Vector Machine
Step 7: Test the images using support vector Machine
Step 8: Calculate the Classification and prediction rate.

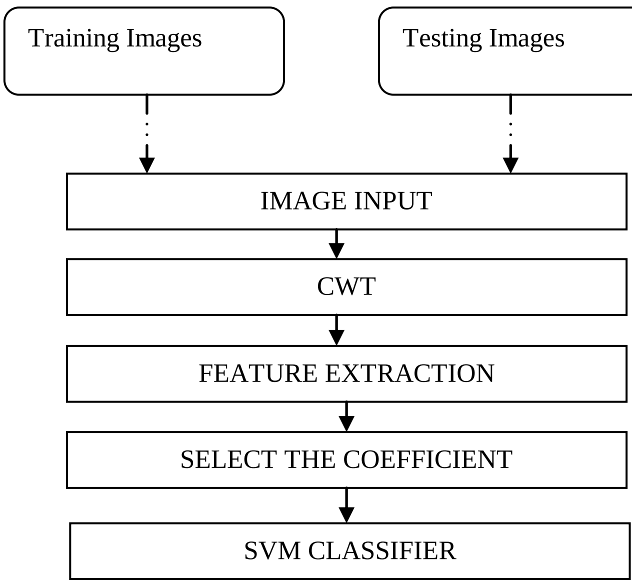


Figure 2. A Block Diagram for the Proposed Methods

3. Feature Extraction Using DWT, RADON, and CWT:

The feature extraction method that we adopted is Continuous Wavelet Transform (CWT). In This Novel based Method, 1-D continuous wavelet transform allows an input image to be decomposed into a set of independent coefficients corresponding to each one dimensional wavelet basis. We use continuous wavelets to make no redundancy in the information represented by the wavelet coefficients, which leads to efficient representation. Also, it provides exact reconstruction of the original image. We also use other wavelet transform such as Discrete Wavelet Transform and Radon Transform. Wavelet coefficient [8][9] represents the "degree of correlation" (or similarity) between the image and the mother wavelet at the particular scale and translation. Thus, the set of all wavelet coefficients [10] gives the wavelet domain representation of the image. After decomposition of the image, the details coefficients can be threshold.

3.1 The Discrete Wavelet Transform (DWT):

A Discrete Wavelet Transform has played a significant role to reduce the dimension of an image and extract the features by decomposing an image in frequency domain into sub-bands at different scales. The DWT of an image is created as follows: In the first level of decomposition, the image is split into four sub-bands, namely HH1, HL1, LH1, and LL1, as shown in Figure 1. The HH1, HL1 and LH1 sub-bands represent the diagonal details, horizontal features and vertical structures of the image, respectively. The LL1 sub-band is the low resolution residual consisting of low frequency components and it is this sub-band which is further split at higher levels of decomposition [14].

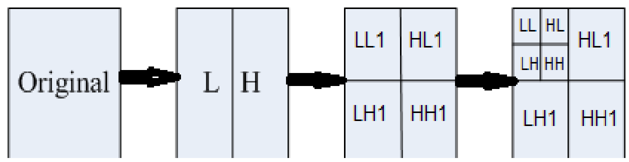


Figure 3: The process of decomposing an

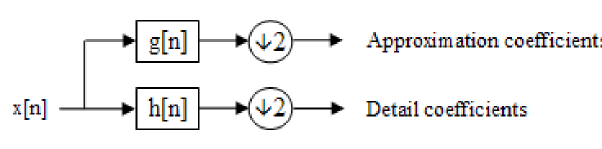
The DWT of a signal x is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response g resulting in a convolution of the two heading

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k] g[n-k].$$

The signal is also decomposed simultaneously using a high-pass filter h. The outputs giving the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as a quadrature mirror filter. However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter outputs are then sub sampled by 2 (Mallat's and the common notation is the opposite, g- high pass and h- low pass):

$$y_{\mathrm{low}}[n] = \sum_{k=-\infty}^{\infty} x[k] g[2n-k]$$

This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled.



With the sub sampling operator $\downarrow$

$$(y \downarrow k)[n] = y[kn]$$

The summation can be written more concisely.

$$y_{low} = (x * g) \downarrow 2$$

$$y_{high} = (x * h) \downarrow 2$$

3.2 The Radon Transform:

The radon transform [15] compute projections of an images matrix along specified direction .Radon transform on an image f (x,y) for a given set of angles can be thought of as computing the projection of the image along the given angles. The resulting projection is the sum of the intensities of the pixels in each direction, i.e. a line integral. The result is a new image R(r,q).should start at the left-hand margin on a separate line.
the Radon transform can be written as

$$R(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(\rho - x\cos\theta - y\sin\theta)dxdy$$

Where the Dirac delta function is is the perpendicular distance of a line from the origin and is the angle formed by the distance vector .

3.3 Continuous Wavelet Transform (CWT):

A continuous wavelet transform (CWT) is used to divide a continuous-time function into wavelet. The CWT has the ability to decompose complex information and patterns into elementary forms. The continuous wavelet transform possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization. The continuous wavelet transform of a continuous, square - integreable function x (t) at a scale a > 0 and translational value b is expressed by the following integral: top of this paragraph illustrates a sub-subheading.

$$X_W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right) dt$$

Where, ψ(t) is a continuous function in both the time domain and the frequency domain called the mother wavelet and represents operation of complex conjugate. The main purpose of the mother wavelet is to provide a source function to generate the daughter wavelets which are simply the translated and scaled versions of the mother wavelet. To recover the original signal x(t), inverse continuous wavelet transform can be exploited.

$$x(t) = \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2}(a,b)\frac{1}{\sqrt{|(a)|}}\tilde{\psi}\left(\frac{t-b}{a}\right) db\, da$$

Where, is the dual function of ψ(t). And the dual function should satisfy

$$\int_0^{\infty} \int_{-\infty}^{1} \frac{1}{|a^3|}\ \psi\left(\frac{t1-b}{a}\right) \tilde{\psi}\left(\frac{t-b}{a}\right) db\ da\ \delta(t - t1)$$

Sometimes, $\tilde{\psi}(t) = C_{\psi}^{-1}\,\psi(t), where,$

$$C_{\psi} = \frac{1}{2}\int_{-\infty}^{+\infty} \frac{|\psi(\zeta)|^2}{|\zeta|} d\zeta$$

is called the admissibility constant and is the Fourier transform of ψ. For a successful inverse transform, the admissibility constant has to satisfy the admissibility condition:

$$0 < c_{\psi} < +\infty$$

It is possible to show that the admissibility condition implies that , so that a wavelet must integrate to zero [11].

The advantage of using wavelet-based coding in image compression is that it provides significant improvements in picture quality at higher compression ratios over conventional techniques. Since wavelet transform has the ability to decompose complex information and patterns into elementary forms, it is commonly used in acoustics processing and pattern recognition. Edge and corner detection, partial differential equation solving, transient detection, filter design, Electrocardiogram (ECG) analysis, texture analysis and business information analysis. Continuous Wavelet Transform (CWT) is very efficient in determining the damping ratio of oscillating signals (e.g. identification of damping in dynamical systems). CWT is also very resistant to the noise in the signal.

4. Classification and Prediction

4.1 Support Vector Machine (SVM):

Consider the pattern classifier, which uses a hyper plane to separate two classes of patterns based on given examples {x (i), y (i)} i=1 n .Where (i) is a vector in the input space I=Rk and y (i) denotes the class index taking value 1 or 0.A support vector machine is a machine learning method that classifies binary classes by finding and using a class boundary the hyper plane maximizing the margin in the given training data [7][12][13]. The training data samples along the hyper planes near the class boundary are called support vectors, and the margin is the distance between the support vectors and the class boundary hyper planes. The SVM [6] are based on the concept of decision planes that

define decision boundaries. A decision plane is one that separates between assets of objects having different class memberships. SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consists of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features).
Given a training set of instance label pairs (xi, yi), i=1… l where xi € R and y € (1,-1)l, the SVM requires the solution of the following optimization problem.

$$\text{Min}_{w, b, \pounds} 1/2^{w}T^{w} + c\ \Sigma^{l}_{i=1}\ \varepsilon_{I}$$

Subject to $yi\ (w^{T}\ Ø\ (xi) + b) > 1 - \varepsilon I,$

$\varepsilon_i >= 0$

Here training vectors xi are mapped into a higher dimensional space by the function Ø. Then SVM finds a linear separating hyper plane with the maximal margin in this higher dimensional space>0.is a penalty parameter of the error term. Furthermore, k (xi, xj) =Ø (xi) Ø (xj) is called the kernel functions.

There are number of kernels that can be used in SVM models. These include linear polynomial, RBF and sigmoid.

polynomial, RBF and sigmoid

$Ø = \{x_i * x_j$	linear
$(\gamma x_i x_j + \text{coeff})^d$	polynomial
$\text{Exp}\ (-\gamma \lvert x_i - x_j \rvert^2)$	RBF
$\text{Tanh}\ (\gamma x_i x_j + \text{coeff})$	sigmoid }

The RBF is therefore the most popular choice of kernel types used in SVM. There is a close relationship between SVMs and the Radial Basis Function (RBF) classifiers. In the field of medical imaging the relevant application of SVMs is in breast cancer diagnosis. The SVM is the maximum margin hyper plane that lies in some space. The original SVM is a linear classifier. For SVMs, using the kernel trick makes the maximum margin hyper plane fit in a feature space. The feature space is a non linear map from the original input space, usually of much higher dimensionality than the original input space. In this way, non linear SVMs can be created. Support vector machinesare an innovative approach to constructing learning machines that minimize the generalization error. They are constructed by locating a set of planes that separate two or more classes of data. By construction of these planes, the SVM discovers the boundaries between the input classes [12]; the elements of the input data that define these boundaries are called support vectors.
For Gaussian radial basis function:

$$K\ (x, x') = \exp\ (-|x-x'|2/\ (2\sigma 2)).$$

The kernel is then modified in data dependent way by using the obtained support vectors. The modified kernel is used to get the final classifier.

5. Results and Discussion

The experiment is carried out with ORL database containing 400 images of male and female .This images are frontal with variation in pose , expression , and various illumines condition . Here 400 images is used for training (350 male , 50 female) ,similarly 200 image (150 male and 50 female) is used for Testing .

In this paper we examine the effects of different feature extraction methods like DWT , Radon , CWT on the face gender classification method SVM. First we compute and select the limited variance of DWT , Radon Coefficients and feed them as inputs to SVM classifier with various Kernels like Linear ,Rbf ,Polynomial , Quadratic , and MLP. Then we compare our approach (the Continous Wavelet Transform with SVM along with Various Kernals) with DWT + SVM and RADON with SVM and the results of error rate are shown in Table 1, 2 and 3.

Table1: Error and Classification rate of DWT w/ SVM

DWT+SVM	Error Rate			
Classifier	**Over All**	**Male**	**Female**	**Classification Rate**
svm w/Linear Kernel	37.75	0	98	75.5
svm w/RBF Kernel	25	0	100	75
svm w/Poly .Kernel	25	0	100	75
svm w/Quardratic Kernel	25	0	100	75
svm w/MLP Kernel	9	11.33	3	91

Table2: Error and Classification rate of Radon w/ SVM

Radon+SVM	Error Rate			
Classifier	**Over All**	**Male**	**Female**	**Classification Rate**
svm w/Linear Kernel	21	0	84	79
svm w/RBF Kernel	24.5	0	98	75.5
svm w/Poly .Kernel	25	0	100	75
svm w/Quardratic Kernel	25	0	100	75
svm w/MLP Kernel	12	12.6	18	88

Table3: Error and Classification rate of CWT w/ SVM

CWT +SVM	Error Rate			
Classifier	**Over All**	**Male**	**Female**	**Classification Rate**
svm w/Linear Kernel	2	0	8	98
svm w/RBF Kernel	18	0	72	82
svm w/Poly .Kernel	25	0	100	75
svm w/Quardratic Kernel	22.5	0	90	77.5
svm w/MLP Kernel	14	11.33	22	86

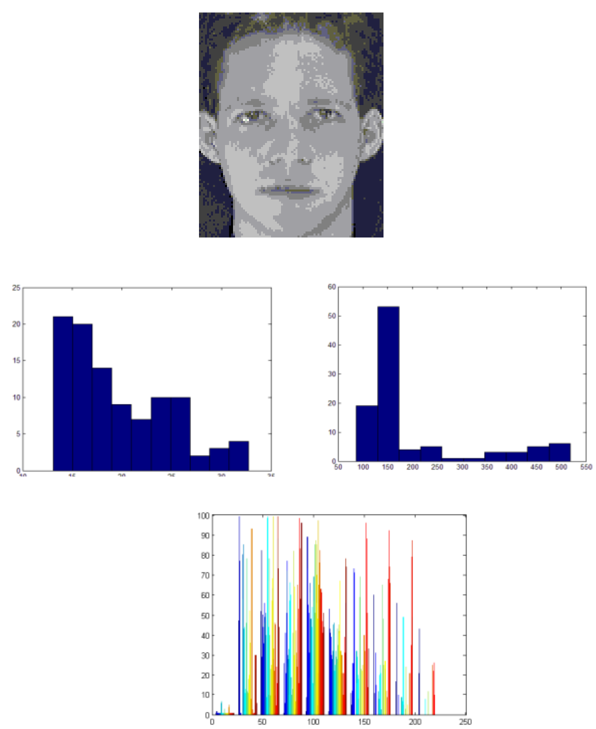


Figure 4. Male Gender Coefficient of CWT, DWT, RADON

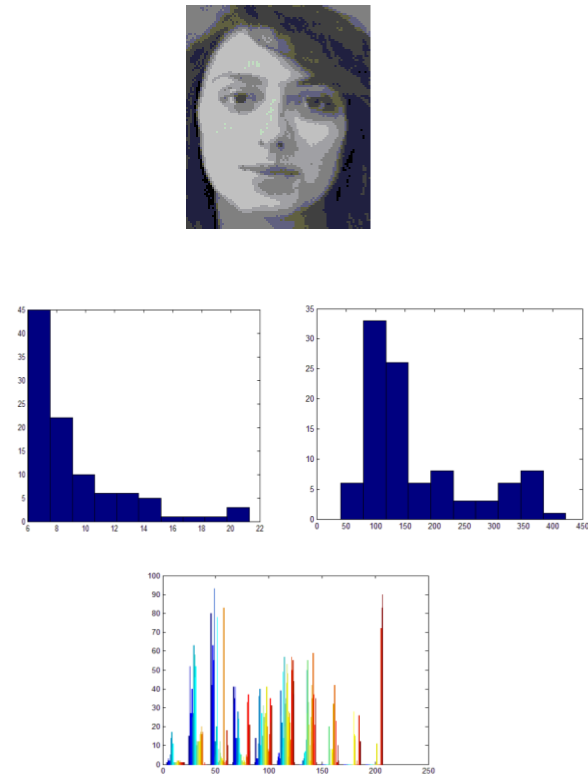


Figure 5. Female Gender Coefficient of CWT ,DWT, RADON

In Figure 4 shows the Male Gender coefficient of continuous wavelet Transform, Discrete wavelet Transform, Radon Transform. Similarly Figure 5 shows the Female Gender Coefficient continuous wavelet Transform, Discrete wavelet Transform, Radon Transform.

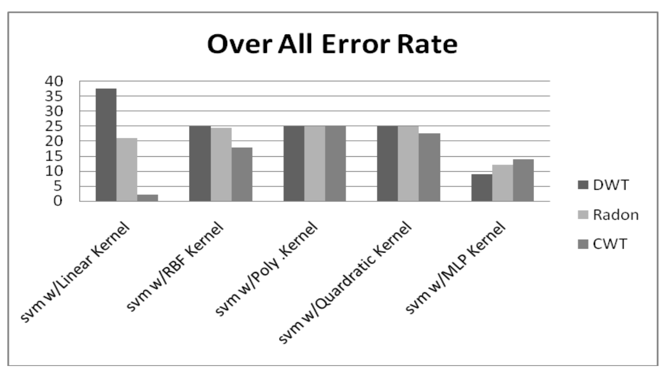


Figure 6. comparison graph of overall error rate of various Transform and SVM with various Kernels .

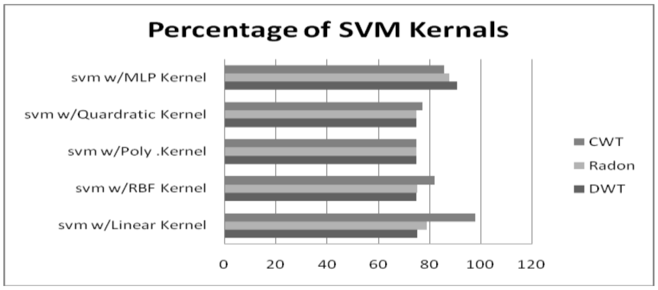


Figure 7. Comparison graph of prediction rate of various Transform and SVM with various Kernels .

In Figure 6 shows the overall error rate of various wavelets transform and support vector machine with different types of kernels. Similarly Figure 7 shows the prediction rate of various Transform and Support vector Machine with various Kernels.

Authors	Methods	Percentage
Tamura et al. (1996) [1].	neural networks	93%
Wiskottet al. (1997) [1].	genetic algorithms(GA)	95.3%.
Costen et al. (2004)[1].	SVM	94.42%
Sun et al. (2006)[1].	(SOM)	95.75%,
Lian and Lu (2006)[1].	SVM	96.75%
Baluja and Rowley (2007)[1].	Adaboost classifier	93%
The Proposed Method	CWT (1-D), SVM	98%

Table 4. Comparison between the existing methods and the proposed

In Table 4, we present the comparison of techniques, authors to achieve the same goal as this paper. The classification rate of the various techniques are Tamura et al. (1996) is 93 % Using neural networks ,Wiskott et al(1997)is 95.3% using genetic algorithms, Jain and Huang shows the percentage of 99.3% using ICA/LDA. Costene et al.(2002) shows the percentage of 94.42% using SVM, Sun et al.(2006) shows the percentage of 95.75 in SOM. Lian and Lu (2006) show the percentage of 96.75 Using SVM. Baluja and Rowley (2007) show the percentage of 93% using Adaboost Classifier.

Here all the experiments are carried out with grey scale, low resolution images unless otherwise specified with fivefold cross validation. After applying numerous Kernels in SVM Classification along with Wavelet Transforms like Discrete wavelet Transform, Radon Transform, continuous wavelet transform. We concluded in our paper SVM with Linear Kernel perform the best along with CWT, which have only 2% error rate and high percent classification accuracy (98%), followed by MLP Kernel which shows the better result and less error rate .More over DWT, Radon when classified with SVM -MLP kernel also shows the better result. Note that error rate occur more in female compare to male, this is due to less number of female images taken into training set.

The classification rate of our technique is superior to the rest up to techniques. Furthermore there is only 2% error rate in our proposed technique. Overall speaking, the proposed novel technique outperforms other techniques in terms of specificity and sensitivity.

6. Conclusions

An original analysis of algorithm to classifying the gender of male and female is distinguish and computed and verified. The images contain variations in lighting and facial expression, pose angles, aging effects etc and finding out the wavelet coefficient using Continuous wavelet transform in gender facial images. CWT is constructed to define the feature of the face gender and best when compared with DWT and Radon. Support Vector Classifier is developed for classification with various Kernels like. Linear, polynomial, RBF and sigmoid. Where we find Linear Kernel is the best choice for our proposed method, which is the original SVM [13]

After Applying CWT, The resulted Coefficient is the binary data: 0 for Male Gender and 1 for Female Gender. The Support Vector Machine classifier with linear kernel exhibits superior efficacy. These algorithms were all coded in MATLAB. Computation time is very less for CWT w/SVM when compared with other techniques likes DWT w/SVM and Radon w/SVM .Computational time depends on the configuration of the PC we used. Therefore, it is better to take the ratio of the computational times into account rather than exact values. Finally we concluded that classification rate is raised up to 98% and there is 2% error rate .if CWT and SVM with Linear Kernel are both employed. Finally, SVM shows the lowest MSE in Classification and Prediction.

In Future, We have to develop the Method to classification the face gender and to adjust or identify the Threshold automatically using a Novel Clustering Algorithms.

References

[1] A. [1] Michael C. Mangini, Irving Biederman. Making the ineffable explicit: estimating the information employed for face classifications. Cognitive Science 28 (2004) 209–226.

[2] Jain, A., Huang, J., May 2004. Integrating independent components and linear discriminate analysis for gender classification. In: Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'04), pp. 159–163.

[3] Saatci, Y., Town, C., April 2006. Cascaded classification of gender and facial expression using active appearance models. In: Proc. 7th Internat. Conf. on Automatic Face and Gesture Recognition (FGR'06), pp. 393–400.

[4] Erno Mäkinen, Roope Raisamo. An experimental comparison of gender classification methods. Pattern Recognition Letters 29 (2008) 1544–1556.

[5] I. Daubechies, "Ten Lectures on Wavelet, "SIAM,pp. 167-213, Philadelphia 1992.

[6] Y. Liu, Y. F. Zhung, "FS_SFS: A novel feature selection method for support vector machines", pattern recognition New York, vol.39, pg.1333-1345, December 2006.

[7].Smola A. J., Scholkopf B., and Muller K. R., "The connection between regularization operators and support vector kernels", Neural Networks New York, vol.11, pg 637-649, November 1998.

[8].Chen HW. Image processing of micro-calcifications for early-stage breast cancer via wavelet analysis and neural network. M.S. Thesis. Department of Mechanical Engineering, National Cheng-Kung University, Tainan, Taiwan, ROC; 2008.

[9]en.wikipedia.org/wiki/Wavelet

[10].en.wikipedia.org/wiki/Wavelet transform

[11].en.wikipedia.org/wiki/Continuous_wavelet_transform.

[12].en.wikipedia.org/wiki/Support Vector Machine

[13] Fahim Mannan, 260 266 294, School of Computer Science, McGill University" Classification of Face Images Based on Gender using Dimensionality Reduction Techniques and SVM."

[14]http://research.ijcaonline.org/volume34/number2/pxc3875859.pdf

[15]http://en.wikipedia.org/wiki/Radon_transform

Amjath Fareeth is a member of IEEE; he received his Master of Computer Applications from Bharathidasan University, INDIA. Pursuing Ph.D. in Karpagam University, INDIA. He has worked at Infotech Global for McKinsey & Company, USA as an external resource. Currently he is working in Taif University as Lecturer. His area of interest is Software Testing, Image Processing, and Bioinformatics.

Gul Shaira Banu received her Masters of Engineering in Computer Science from Anna University, INDIA. Pursuing Ph.D. in Karpagam University, INDIA. She has worked as Lecturer in PSR Engineering College Tamil Nadu, India. Currently working as a Lecturer in College of Computers and Information Technology, Taif University, Kingdom of Saudi Arabia. Her research interests are Machine Learning, Bioinformatics and Medical Image Processing.