Performance Improvements in Face Classification using Random Forest

Vatsal Vishwakarma*, Abhishek Kumar Srivastava **
*(Department of Electronics and Communication, Lovely Professional University, Jalandhar , India.)
** (Department of Electronics and Communication, Lovely Professional University, Jalandhar , India.)

ABSTRACT

Face classification can be defined as the problem of assigning a predefined label to an image or subpart of an image that contains one or more faces. This definition comprises many sub disciplines in the visual pattern recognition field: (i) face detection, where the goal is to detect the presence of a face on an image, (ii) face recognition, where we assign an identifier label to the detected face, Face recognition is emerging as an active research area with numerous commercial and law enforcement applications. Although existing methods perform well under certain conditions, the illumination changes, occlusions and recognition time are still challenging problems. This work attempts to use Random Forests to deal with the above challenges in improvement in face classification. Random Forest is a tree based classifier that consists of many decision trees. Each tree gives a classification[11] and regression[11] and process further carry with this two techniques ,here we are focusing main over the regression technique .The proposed algorithm first extracts features from the face images from a small dataset using the Gabor wavelet transform and then uses the Random Forest algorithm to classify the images based on the regression technique. The proposed algorithm makes use of a Random Forest regression that selects a small set of most discriminant Gabor wavelet features. Only this small set of features is now used to classify the images resulting in a fast face recognition technique. The proposed approaches are tested on a multiple image face databases and the results are found to be highly encouraging.

Keywords **–** Gabor Wavelet, Face Classification, Random Forest ,Regression

1. INTRODUCTION

Machine recognition of faces is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision and neural networks. Face recognition technology has numerous commercial and law enforcement applications. These applications range from static matching of controlled format photographs such as passports, credit cards, photo ID0s, drivers license0s, and mug shots to real time matching of surveillance video images. Understanding the human mechanisms employed to recognize faces constitute a challenge for psychologists and neural scientists. In addition to the cognitive aspects, understanding face recognition is important, since the same underlying mechanisms could be used to build a system for the automatic identification of faces by machine. Although, humans seem to recognize faces in cluttered scenes with relative ease, having the ability to identify distorted images, coarsely quantized images, and faces with occluded details, machine recognition is much more daunting task. A formal method of classifying faces has been first proposed by Francis Galton . Research interest in face recognition has grown significantly in recent years as a result of the following facts:

1. The increase in emphasis on civilian/commercial research projects,
2. The increasing need for surveillance related applications due to drug trafficking, terrorist activities, etc.
3. The re-emergence of neural network classifiers with emphasis on real time computation and adaptation,
4. The availability of real time hardware, Even now, most of the access control methods, with all their legitimate applications in an expanding society, have a bothersome drawback. Except for human and voice recognition, these methods require the user to remember a password, to enter a PIN code, to carry a badge, or, in general, require a human action in the course of identification or authentication. In addition, the corresponding means (keys, badges, passwords, PIN codes) are prone to being lost or forgotten, whereas fingerprints and retina scans suffer from low user acceptance. Modern face recognition has reached an identification rate greater than 95% with well-controlled pose and illumination conditions. While this is a high rate for face recognition, it is not comparable to methods using keys, passwords or badges.

2. METHODOLOGY

2.1 Gabor Wavelet

Gabor wavelets form an excellent filter for spatial localization as well as orientation selection. Moreover, they

Are very robust against variations due to illumination and Changes in facial expressions. A Gabor Wavelet ψu,v(z) is Defined as [15]:

$$\psi_{u,v}(z) = \frac{||k_{u,v}||^2}{\sigma^2} e^{-\frac{||k_{u,v}||^2 ||z||^2}{2\sigma^2}} \left[e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right] \quad .(1)$$

where *z* = (*x*,*y*) gives the horizontal *x* coordinate and vertical *y* coordinate of the point and the parameters *u* and *v* define the orientation and scale of the Gabor filter while the wave vector ku,v is defined as follows:

$$k_{u,v} = k_v e^{i\phi_u} \quad ..(2)$$

where kv = kmaxfv gives the scale and φu = πu 8 gives the orientation. In this work 5 different scales, v _ {0, ..., 4} and 8 different orientations, u _ {0, ..., 7} are chosen. Thus the Gabor wavelet representation Ru,v(z) is the convolution of the image I(z) with the family of 40 Gabor kernels ψu,v(z). The output Ru,v(z) of each Gabor kernel is a complex function, so the magnitude response ||Ru,v(z)|| is used to represent the features. Therefore, a Gabor wavelet feature *j* is configured by the three key parameters: position *z*, orientation *u* and scale *v*.

$$j(u,v,z) = ||R_{u,v}(z)|| \quad (3)$$

The result of convolving a face image with the Gabor kernels is shown in Figure 1. If we have an image of size 50x50 then convolution with Gabor filters would lead to 50x50x40 = 100000 features. High feature space dimension is the major disadvantage of using Gabor wavelets. Generation of these many features for each probe image takes large amount of time. We believe that a large number of redundant features are generated by the Gabor filters. There is a need to discard all redundant features generated by the filter. This paper proposes a technique based on Random Forests to select the most discriminant Gabor features for face recognition.

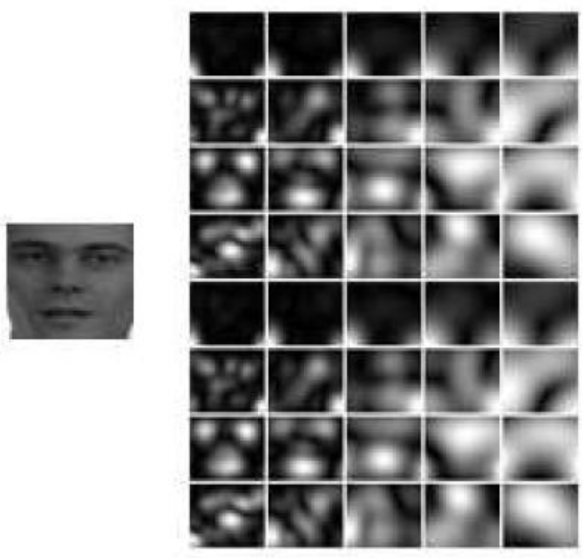

Fig 1. Gabor Representation of a face in general

2.2 Random Forest

A powerful new approach to data exploration, data analysis, and predictive modeling

- Developed by Leo Breiman(father of CART®) at University of California, Berkeley
- Has its roots in CART
- Learning ensembles, committees of experts, combining models
- Bootstrap Aggregation (Bagging)
- CART-Classification Regression Trees

Breiman (2001) proposed random forests, which add an additional layer of randomness to bagging. In addition to constructing each tree using a different bootstrap sample of the data, random forests change how the classification or regression trees are constructed. In standard trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node. This somewhat counterintuitive strategy turns out to perform very well compared to many other classifiers, including discriminant analysis, support vector machines and neural networks, and is robust against overfitting [1]The Random Forest algorithm can be summarized as follows [17]:

- Draw *n*tree bootstrap samples from the original data.
- For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample *m*try of the predictors and choose the best split from among those variables.
- Predict new data by aggregating the predictions of the *n*tree trees (i.e., majority votes for classification and average votes for regression).

Random Forest only uses $2/3^{rd}$ of the training data to build the random Forest model ,remaining $1/3^{rd}$ of the training data can be used to estimate the error of the prediction .Decision trees are predictive model that uses a set of binary rules to calculate the target value.

a) Features of Random Forest [11]

- It is unexcelled in accuracy among current algorithms.
- It runs efficiently on large data bases.
- It can handle thousands of input variables without variable deletion.
- It gives estimates of what variables are important in the classification.
- It generates an internal unbiased estimate of the generalization error as the forest building progresses.
- It has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.
- It has methods for balancing error in class population unbalanced data sets.
- Generated forests can be saved for future use on other data.
- Prototypes are computed that give information about the relation between the variables and the classification.
- It computes proximities between pairs of cases that can be used in clustering, locating outliers, or (by scaling) give interesting views of the data.

- The capabilities of the above can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier detection.
- It offers an experimental method for detecting variable interactions.

a) Trees can be combined By Voting or Averaging

Trees combined via voting (classification) or averaging (regression)[7].

Classification trees "vote"

- Recall that classification trees classify
- assign each case to ONE class only
- With 50 trees, 50 class assignments for each case
- Winner is the class with the most votes
- Votes could be weighted or say by accuracy of individual trees

Regression trees assign a real predicted value for each case

- Predictions are combined via averaging
- Results will be much smoother than from a single tree

b) RF Self Testing

- Each tree is grown on about 63% of the original training data (due to the bootstrap sampling process)
- Left out 37% of the data is available to test any single tree .
- Use this left out data, named "Out of Bag" or OOB to calibrate performance of each tree
- Use OOB data to also keep a running tab on how often each record is classified correctly when it belongs to OOB
- All performance statistics reported by RF are based on OOB calculations

In simple word ,RF only uses 63% of the training data to grow one tree and remaining 37% is used for the OOB error rate.[7]

c) The OOB Error Estimate[7]

- RF does not require use of a test dataset to report accuracy and does not use conventional cross-validation
- For every tree grown, about 37% of data are left out-of-bag (OOB)
- These cases OOB cases are used as test data to evaluate the performance of the current tree
- For any tree in RF, its own OOB sample is used: a true random sample
- The final OOB estimate for the entire RF can be simply obtained by cumulating individual OOB results on a case-by-case basis
- Error rate for a case estimated over subset of trees in which itis OOB
- Error estimate is unbiased and behaves as if we had an independent test sample of the same size as the learn sample

2.3 Random Forest Classification

Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes [5].

Each tree is grown as follows:

1. If the number of cases in the training set is N, sample N cases at random - but *with replacement*, from the original data. This sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

It is often of interest to know which of the variables are important in classification. There are two measures of importance in Random Forests, the mean decrease in accuracy and the Gini index. They give possible ways to quantify which genes are most informative, i.e. contribute most to the prediction accuracy, for giving the correct pathway-based classification.[8]

2.4 Random Forest Regression

The Random Forests regression tree is built in a similar fashion as that in the classification method. The goal for the regression method is to find a tree that best predicts the continuous outcome of the given dataset for research.[8] As in the classification case, two importance measures can be obtained for regression as well. They are the mean decrease in accuracy and mean decrease in MSE.[8] Breiman shows that, while standard decision tree alone suffer from overfitting, a collection of randomly trained trees has high generalization power. Random forests are thus ensembles of trees trained by introducing randomness[5]

2.5 Selection of Regression technique as compared to Classification

The Random Forest algorithm makes no distinction between

the relevance of features during construction of the forest during the regression or Classification[19]. At each node, the features are selected randomly with equal probability. Therefore, the performance can suffer significantly from the presence of redundant features. Basically RF is ensemble classifier based on regression and classification[1]. But RF classification perform poorly in instance of the class imbalance ,this led to the introduction of the additional class weighing parameter that will make the classification calculation complicated ,so sometimes loss of accuracy occurs[3],basically Estimation or the prediction of the unknown values of one variable from known value is known as the Regression simply says IF two values are significantly co-related ,so it possible to predict values of one variable from other & this is known as Regression analysis .Some advantages of Regression Analysis:

- Regression tree is effective in uncovering structure in data with hierarchical and non-adaptive variable as compared with classical statically method(classification)[4].
- Primary disadvantage of Boosting trees is that it requires minimum 30-80 trees ,so interpreting multiple individual trees become impossible.[4]
- Boosting mainly used over large dataset as to classify them using classification .[4]
- Basically ,regression trees are grown without pruning and averaged ,the variance component of the output error reduced (Briemann,1996).

So doing work on MATLAB platform ,regression leads in some way ,specially on small datasets .Also ,RF is a effective tool in prediction specially on large dataset they do not over fit, right kind of randomness make them accurate classifier and regressor.(Leo D Brriemann).

Following steps were done to acquire the images and analysis further;
1 .Take the image database of different faces
2. RF regression
3. Random forest selects the most discriminate features by averaging
4. Bagging for regression.
5. Select the best feature for face classification

2. EXPERIMENT RESULT

We have taken Extended Yale database with cropped images[9] to classify it with the RF Regression method ,using of Gabor filter and getting features also has been done for the research work but takes a bit time on MATLAB to classify ,so showing the experiment result with the help of another database. Main purpose of the paper is to show the regression results on MATLAB platform

Take a image database(The Extended Yale database B) on MATLAB platform[9]

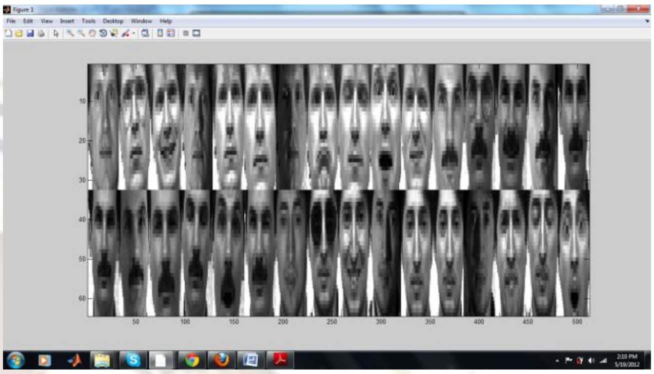

Fig 2: To display the Face image of the database

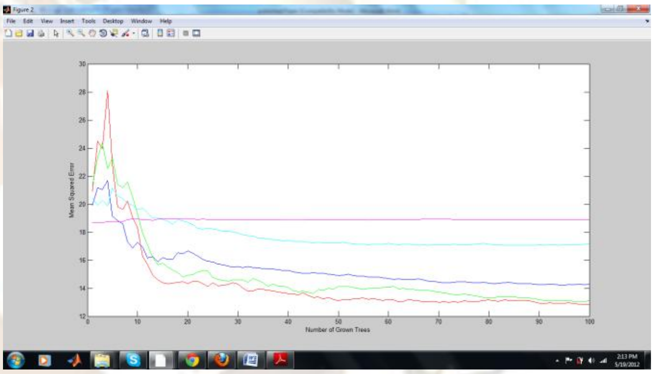

Fig 3: Graph plot between no. of grown trees and MSE (Mean sq. Error)

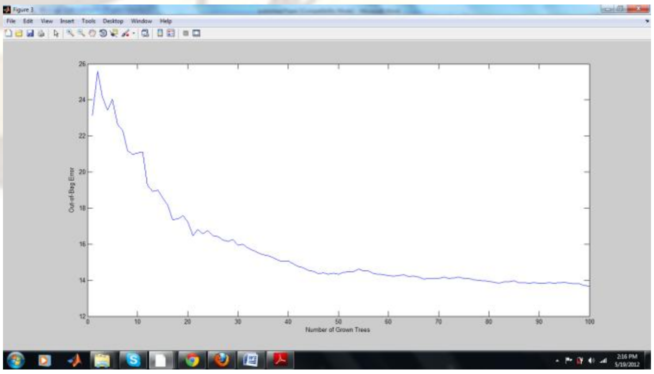

Fig 4: Graph plot b/w no of grown trees and OOB error rate

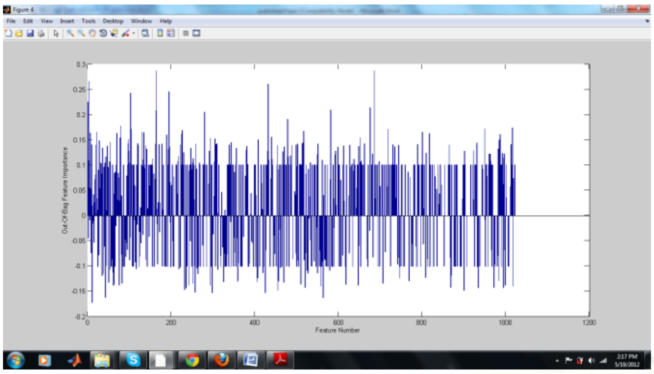

Fig 5: Graph plot b/w Feature num and OOB feature importance

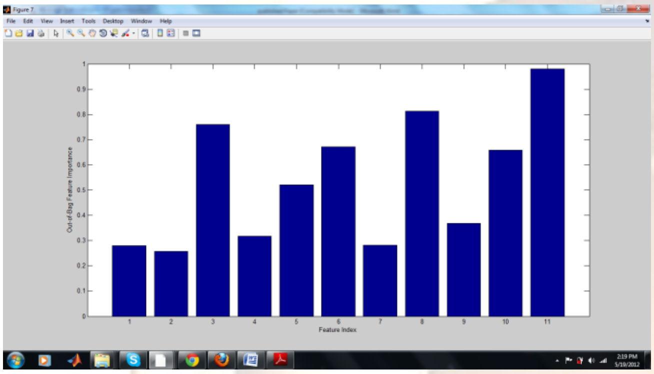

Fig 6: Graph plot b/w OOB feature importance and Feature index

4. CONCLUSION

In the past few years, tree based classifiers and ensemble learning techniques have attracted attention and achieved success as suitable classifiers for face recognition. Random forests is one of the best classification and regression technique available today and has been shown to perform very well compared to other classifiers, including discriminant analysis, support vector machines and neural networks. The advantage of using Regression technique is that it apply easily on small dataset and can easily handle on MATLAB platform ,so ultimately do not require any other special software like R (ver 9),Statica ,Nimbus tree ,Apache or any other for Random forest classifier. Gabor features are already a proven success in the face recognition area. The advantages of Gabor features include their localizability, orientation selectivity, and spatial frequency characteristics. The aim of this work is to bring together the best of both feature extraction and regression techniques. Gabor features of a face image are classified using random forests. However, any technique using all Gabor features suffers from the problem of high dimensionality and is not suitable for smooth usage in a practical scenario. This work tackles this problem by selecting the most discriminant gabor features and using only these features in the Face classification process.. The experiments on subset of the Face image database have shown that these hundreds of Gabor features are enough to achieve good performance compared to using the complete set of Gabor features. This shows the effectiveness of the proposed method with help of Regression technique.

REFERENCES

[1] Classification and Regression by Random Forest by Andy liaw ,Mattew weiner ,dec(2002)

[2] “Face recognition homepage.” [Online]. Available: http://www.face-rec.org/algorithms/

[3] Machine learning benchmarks and RF Regression by MR segal (2004)

[4] Classification and Regression tree techniques (2007).

[5] breiman/RandomForests/cc_home.htm

[6] Face classification using Random Forest by vidhut ghoshaal ,paras Tikmani (2009).

[7] Data mining with Random Forest by Dan steinberg,Mikhai Golovanya,N.scott Cardell (2004)

[8] Pathway analysis using random forests classification and regression(2006) By Herbert Pang1, Aiping Lin2, Matthew Holford1, Bradley E. Enerson3, Bin Lu5, Michael P. Lawton5, Eugenia Floyd5 and Hongyu Zhao1,2,4,

[9] Acquiring Linear Subspaces for Face Recognition under Variable Lighting. By Kuang-Chih Lee, Student Member, IEEE, Jeffrey Ho, Member, IEEE, and David Kriegman, Senior Member, IEEE(may,2005)

[10] E. Bauer and R. Kohavi, “An empirical comparison of voting classification algorithms: Bagging, boosting, and variants,” in *Machine Learning*, vol. 36, no. 1/2, 1999, pp. 105–139.

[11] L. Breiman, “Random forests,” in *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32.

[12] S. N. A.Z. Kouzani and K. Khoshmanesh, “Face classification by a random forest,” in *Proceedings of TENCON 2007 - 2007 IEEE Region 10 Conference*, 2007.

[13] T. D. Vaishak Belle and S. Schiffer, “Randomized trees for real-time one-step face detection and recognition,” in *Proceedings of InternationalConference on Pattern Recognition 2008*, 2008.

[14] I. R. Fasel, M. S. Bartlett, and J. R.Movellan, “A comparison of gabor filter methods for automatic detection of facial landmarks,” in *Proceedings of the 5th International Conference on Face and Gesture Recognition*, 2002, pp. 231–235.