Attribute Based Face Classification Using Support Vector Machine

Brindha.M[1], Amsaveni.R[2]

[1]*Research Scholar, Dept. of Computer Science, PSGR Krishnammal College for Women, Coimbatore*
[2]*Assistant Professor, Dept. of Information Technology, PSGR Krishnammal College for Women, Coimbatore.*

---***---

Abstract - Verifying human face from captured images of his/her face is challenging task due to various poses and camera focus. Generally, existing approaches to this problem use the features of pose and expression for human face. In this work, different appearance features are taken into concern for face verification. The face verification uses ten different personalities dataset with 61 different attributes of 20 different images for each person using dissimilar poses and expressions. There are two different classification approaches are used for face verification. First approach is attribute based classifiers, it uses binary classifiers trained to identify the presence or absence of describable features of visual appearance like gender, age, skin and, etc. second approach is "face" classifiers, which uses multiclass classifiers trained to verify different persons based on various features. It requires manual labeling because of fewer reasons like similarity of faces, regions of faces to specific reference people. The dataset used in this research work is real-world images of public figures such as celebrities and politicians obtained from the internet. Support Vector Machine (SVM) is used as both binary classification and multiclass classification for attribute classification and face verification respectively. There are eight different evaluation measures are used in face verification system. And the subset evaluation method is used to select attributes from the dataset. Selected attributes are used for classification based on their presence and absence of the visual appearance features. The experiment result shows that proposed work gives 98% of accuracy in face verification and various accuracy results for attribute classification.

Key Words: SVM, face verification, LFW, RBF, FERET, and SimMSVM, etc,.

1. INTRODUCTION

There is vast amount of unpredictability in the manner in which the same face presents itself to a camera: not only might the pose differ, but also differ in hairstyle and expression [1]. In this field of research the illumination direction, focus, resolution, image compression and camera type are almost differ to make face recognition and pattern matching is a challenging task [2], [3]. These various differences in images of the same person have problem of confounded methods for face recognition and verification, often limiting the consistency of automatic algorithms to the domain of more proscribed settings with supportive focus [4].

In recent times, there has been major work on the Labeled Faces in the Wild (LFW) data set for face classification [5]. This dataset is significant in its variability, showing all of the dissimilarities mentioned over. Not astonishingly, LFW has demonstrated the difficult for automatic face verification methods [6], [7]. When one examines the failure cases for some of the previous methods, lots of mistakes are found such as men being confused for women, young people for old, etc. On the other hand, small changes in pose, expression, or lighting can cause the same person to be misclassified by an algorithm as different. In these reasons, researchers have shown interest in the field of face verification systems.

The rest of this paper was prepared as follows; section II describes about previous work related to face verification and section III illustrates the methodology used for face verification and attribute based classification. Section IV demonstrates the experiments and its results in detail. Finally section V gives the detail about conclusion and future work.

2. RELATED WORK

Huang et al [8] uses Labeled Faces in the Wild (LFW) benchmark data set and similar data sets used for face verification based on 2D alignment strategies are used for aligning each pair of images. Wolf et al [9] proposed an approach for binary path features to identify individuals in two ways: one is about to recognize describable attributes and another one is about to recognize similarity to set of reference people.

Cottrell et al [10] and Golomb et al [11] presented automatic gender determinations using neural network approach. Moghaddam [12] extended the research work using support vector machines. Shakhnarovich et al [13] proposed approach framework for face detection and classification based on ethnicity. Bartlett et al [14] and Huang et al [15] developed an application for face detection based on pose and expressions.

Gallagher et al [16] estimates the age and gender to compute the likelihood of first names being related with a particular face. M.J. Lyons et al. [17] implemented Principle Component Analysis (PCA) and Linear Discriminate Analysis (LDA) to evaluate the expression training sets, and got correctness of 92% on Japanese Female Facial Expression (JAFFE) database.

C. Padgett et al. [18] trained a back-propagation neural network for face reorganization yeilds average recognition rate of 86% on Ekman's photos dataset. T. Otsuka et al. [19] used hidden Markov model (HMM) based classifiers to recognize one of six different facial expressions on near real time dataset. M.S. Bartlett et al. [20] proposed Gabor feature based AdaSVM method to distinguish expression, and attain a good performance on Cohen-Kanade expression database.

Brunelli and Poggio [21] developed HyperBF Networks for gender classification in which two opposing RBF networks, one is male and the other one is female, are trained using 16 geometric features as inputs. The results on a dataset of 168 images show an average error rate of21%.

Golomb [22] and Cottrell [23], Tamura et al. [24] applied multilayer neural networks to categorize gender from face images with multiple resolutions. The experiments on 30 test images show that network is able to find Gender from face images of 8-by-8 pixels with an average Error rate of7%.

Wiskottetal. [25] used labeled graphs of two-dimensional views to describe faces. They use a small set of controlled model graphs of males and females to train the general face knowledge. It characterizes the face image space and is used to generate graphs of new faces by elastic graph matching. The error rate of this experiment on a gallery with 112 face images is 9.8%.

RecentlyGutta, Wechsler and Phillips [26] proposed a hybrid method which includes ensemble of neural networks (RBFs) and inductive decision trees with C4.5 algorithm. Experimental results on a subset of FERET images of 256-by-384 resulted in an average error rate 4% for gender classification.

In Moghaddam & Yang [27] 256-by-384 FERET "mugshots" were pre-processed and sub sampled to 21-by-12 pixels for very low-resolution experiments. The experiment used a total of 1,755 FERET images with a 5-fold Cross Validation evaluation methodology. The Best error rate reported was 3.4% using nonlinear Support Vector Machines.

3. METHODOLOGY

3.1 Support Vector Machine (SVM)

Machine learning is the method in which an algorithm improves or "learns" through an experience [28]. The support vector machines is to separate the data by using optimal method. There are two different types of machine learning methods one is supervised learning in which a machine learns to separate between groups based on a training set, whereas with unsupervised learning, there is no equivalent target output.

First, consider the case where the data can be separated into two groups by a hyperplane without any training errors. If the data have this property, they are said as linearly separable and hyperplane that separates data is called a "separating hyperplane" [29].

The general form of a hyperplane is $\langle w, x\rangle + b = 0$, and the decision function for a hyperplane $f(x) = \langle w, x\rangle + b$ can be used as a classification rule by assigning an observation to the positive class (y = 1) for f(x) ≥0 and the negative class (y =− 1) otherwise.

The functional margin of an observation is defined as $\gamma_i = y_i(\langle w, x_i\rangle + b)$ The portion in the parentheses is simply the decision function; therefore, γi ≥0 if and only if the true class and the decision function have the same sign (where by convention, sign (0) =1). However, having the same sign implies the decision function correctly classifies xi. Now, suppose the functional margin of the closest point to a separating hyperplane is k.

This can be written in a more compact form as: $y_i(< w, x_i > +b) \geq 1$. That is, the functional margin for all observations is greater than or equal to one. Thus, one can represent the maximization problem as

Maximize γ

Subject to $y_i(< w, x_i > +b) \geq 1$

In order to find a maximal margin hyperplane, $\gamma = 1/\|w\|$ must be maximized, or equivalently, ||w|| must be minimized. Therefore the minimization problem can be written as

Maximize ||w||

Subject to $y_i(< w, x_i > +b) \geq 1$

3.2 Multiclass Support Vector Machine (MSVM)

Currently there are two types of approaches for multi-class SVM. One is by constructing and unites several binary classifiers whereas the other is by directly considering all data in one optimization formulation [30]. Up to now there are still no comparisons which cover most of these methods.

The simplest model of support vector machine is called Maximal Margin classifier, constructs a linear separator given by $\mathrm{w\,T\,x} - \gamma = 0$ between two classes of examples. The free parameters are vector of weights w is orthogonal to the hyperplane and a threshold value . These parameters are attaining by solving the following optimization problem using Lagrangian duality as follows;

$$\text{Minimize } \frac{1}{2}\|W\|^2$$

$$\text{Subject to } D_{ii}(\mathrm{W^T}X_i - \gamma) \geq 1, \mathrm{i} = 1, \ldots, \mathrm{l}.$$

Where D_{ii} equals to class labels, which take for granted value +1 and –1. The data points with non-null weights are referred support vectors.

The over fitting problem is avoided with help of outliers and wrongly classified training samples. A vector of slack variables that measure the amount of violation of the limitation is bring in the optimization problem is referred as soft margin is given below

$$\text{Minimize } \quad \mathrm{C}\sum_{i=1}^{l}\xi\mathrm{i} + \frac{1}{2}\|W\|^2$$

$$\text{Subject to } D_{ii}(\mathrm{W^T}X_i - \gamma) \geq 1, \mathrm{i} = 1, \ldots, \mathrm{l}.$$

$$\xi\mathrm{i} \geq 0$$

The minimization of the objective function affects the maximum separation between two classes with the minimum number of points crossing particular bounding planes. The parameter referred as a regularization parameter that reins the trade-off between the two terms in the objective function [31]. The proper choice of C is critical for good generalization power of the classifier. The following decision rule is used to accurately calculate the class of new instance with a minimum error.

The advantage of the dual formulation is that it allows a competent learning of non–linear SVM separators, by introducing kernel functions. Technically, a kernel function computes a dot product between two vectors and mapped into a high dimensional feature space. As there is no need to perform this mapping explicitly, the training is still possible although the dimension of the real feature space is very high or infinite.

$$f\,(\mathbf{x}) = \mathrm{sgn}[W^T x - \gamma]$$

The parameters are obtained by solving the following nonlinear SVM dual formulation (in Matrix form),

$$\text{Minimize } \mathrm{LD}\,(\mathbf{u}) = \frac{1}{2}u^T Qu - e^T u$$

$$d^t u = 0, \mathbf{0} \leq \boldsymbol{u} \leq \boldsymbol{Ce}$$

Where and is kernel matrix. The kernel function may be polynomial or RBF (Radial Basis Function) is used to construct hyper plane in the feature space, which separates two classes linearly, by performing computations in the input space [32]. The decision function in this nonlinear case is given by

$$f\,(\mathbf{x}) = \mathrm{sgn}\,[\mathrm{K}\,(x, x^T)^* u - \gamma]$$

Where, u refers the Lagrangian multipliers. When number of classes is more than two, then the problem is called multiclass SVM. There are two types of approaches for multiclass SVM [33], [34]. In the first method called indirect method, several binary SVM's are created and the classifier's output are shared for finding the final class. In the second method called direct method. The formulation of one of the direct methods called as Crammer and Singer Method [20] is

$$\text{Minimize } \frac{1}{2}\sum_{k=1}^{N}(\mathrm{w}_k^T)w_k + \mathrm{C}\sum_{i=1}^{n}\xi\mathrm{i}$$

Subject to the constraints

$$\mathrm{w}_{w_i}^T \emptyset(x_i) - \mathrm{w}_k^T \emptyset(x_i) \geq \mathrm{e}_k^t - \xi\mathrm{i}, \forall K \neq K_i$$

Where is the class to which the training data belong,

$$\mathrm{e}_k^i = 1 - \mathrm{C}_k^i$$

$$\mathrm{C}_k^i = \begin{cases} 1 \text{ if } K_i = K \\ 0 \text{ if } K_i \neq \mathrm{K} \end{cases}$$

The decision function for a new input data is given by

$$\bar{\mathrm{d}}_j = \arg\max\{f_k(x_j)\}$$

$$\text{Where } f_k(x_j) = \mathrm{w}_k^T \emptyset(x_j) - \gamma_k$$

This kind of models mixes multiple binary-class optimization problems into one single objective function and concurrently achieves classification of multiple classes. Nevertheless, a larger computational complexity is required the size of resulting Quadratic Programming (QP) problem. The Simplified Multi-class SVM (SimMSVM) gives a direct solution for training multiclass predictors.

4. EXPERIMENTS AND RESULTS

The experiment was carried out by implementing Support Vector Machine (SVM) in R data mining tool.

4.1 Data Preparation

The face verification uses ten different personalities dataset with 61 different attributes of 20 different images for each person using dissimilar poses and expressions. The dataset used in this research work is real world images of public figures such as celebrities and politicians obtained from the internet. The personalities like Bill Clinton, Atal Bihari Vajpayee, Carlos Menem, Colin Powell, David Beckham, Donald Rumsfeld, George Robertson, George W Bush, Gerhard Schroeder and Gloria Macapagal Arroyo.

4.2 Attribute Selection

Attribute selection is performed using weka platform. The subset evaluation method is used for attribute selection. Attribute classification is done based on the selected attributes. 12 different attributes are selected from 61 attributes for different classification in account of face verification. The selected attributes are White, Baby, Black hair, Blond hair, Blurry, Busy eyebrows, Narrow eyes, Pointy nose, Pale skin, Mouth slightly open, fully visible forehead and Smiling. Each attributes are used for classification of each personality.

4.3 Attribute Classification

Build classifiers to detect the describable attributes of faces. Attribute classification as a supervised learning problem. Training needs a set of positive and negative examples for each attribute. The forward feature selection is used to adding each residual feature to the current feature set.

Each attribute classifier is an SVM with matlab implementation. The accuracies for each selected attributes (12 attributes) classifiers are trained using SVM classifier. The classification accuracies are given in below table (Table 1).

Table -1: Attribute Classifiers

Attributes	Classifiers
White	72.42%
Baby	77.24%
Black hair	81.13%
Blond hair	79.11%
Blurry	80.35%
Busy eyebrows	70.26%
Narrow eyes	84.73%
Pointy nose	85.82%
Pale skin	86.47%
Mouth slightly open	64.49%
Fully visible forehead	93.10%
Smiling	95.33%

4.4 Face Classification

All the experiment estimation performance on face verification task of given faces dataset and determines if the result show the same individual. The face verification is performed on 200 pairs of images of 10 people, divided into 10 cross-validation folds with mutually disjoint sets of each pair of people. Support Vector Machine (SVM) is used as classifier for human face classification problem. Classifier is evaluated based on eight different measures shown in the subsequent table. The following table (Table2) gives the factors of evaluation measures like sensitivity, specificity positive prediction value, detection rate, negative prediction value, prevalence, detection prevalence, kappa and overall accuracy. SVM perform well in face classification with accuracy of about 98.50%.

Table -2: Evaluation Measures

Evaluation Measures	Prediction Rate
Sensitivity	0.9851

Specificity	0.9983
Positive Prediction Value	0.9860
Negative Prediction Value	0.9936
Prevalence	0.1
Detection Rate	0.0985
Detection Prevalence	0.1
Kappa	0.9833
Over all Accuracy	*0.9850 (98.50%)*

The following figures (Chart1 & Chart2) shows the chart representation of evaluation measures. Fig1 illustrates the chart form of first five measures as given in the above table. Fig2 demonstrates the next four different measures and the linear prediction rate of accuracy with R-Squared value of 0.700.

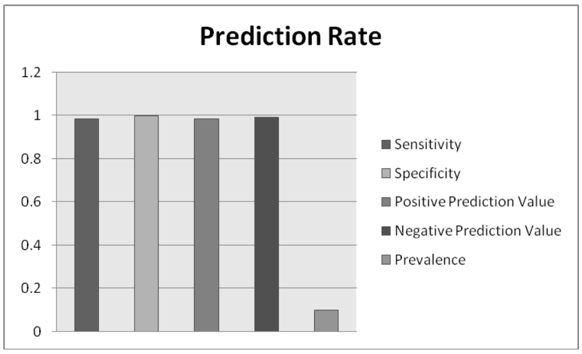

Chart -1: Prediction Rate Comparison for first five measures

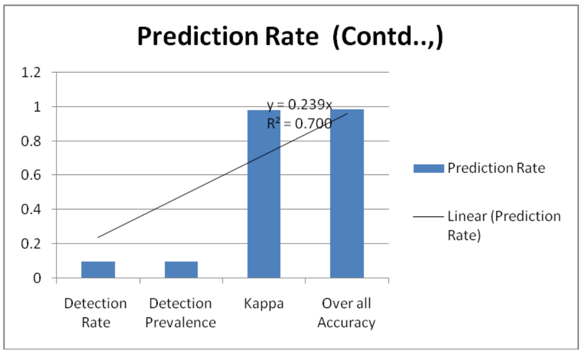

Chart -2: Prediction Rate Comparison for Remaining Measures

The table (Table3) gives of balanced accuracy for 10 dfiffernt classes. Each class refers the each personalites like Bill Clinton, Atal Bihari Vajpayee, Carlos Menem, Colin Powell, David Beckham, Donald Rumsfeld, George Robertson, George W Bush, Gerhard Schroeder and Gloria Macapagal Arroyo. The class 1, 2, 3, 7 and 10 provide the balanced accuracy of 100% for face classification. Class 4 balanced accuracy is 0.9737; class 5 is 0.9944 and so on.

Table -3: Balanced Accuracy for each Class

Class Name	Balanced Accuracy
Class 1-3, 7 and 10	1.0000
Class 4	0.9737
Class 5	0.9944
Class 6	0.9750
Class 8	0.9773
Class 9	0.9973

Figure (Fig1) shows the classifier result of face classification using Support Vector Machine (SVM). The classifier was implemented in R-Data mining tool.

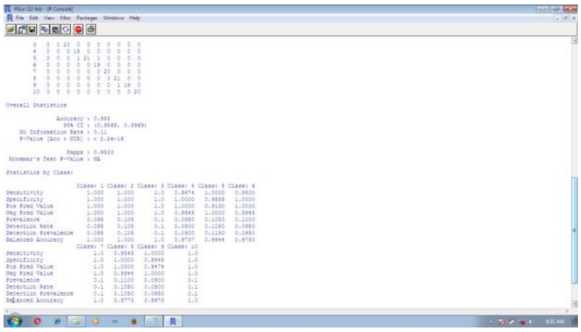

Fig -1: Result for SVM classifier

5. CONCLUSIONS

This research work presented and evaluated two methods for face verification like attribute classification and face classification. In this work Support Vector Machine (SVM) is used as binary classifier and mutli-class classifier in the task of face verification. The dataset contains 10 different people with different poses and expressions; 20 records for each peoples with 61 different attributes. It requires manual labeling because of fewer reasons like similarity of faces, regions of faces to specific reference people. The dataset used in this research work is real-world images of public figures such as celebrities and politicians obtained from the internet. Subset evaluation method with bestfirst function is used for attribute selection, based on the method 12 attributes are selected for classification task. For each people classification accuracy is evaluated by using evaluation measures. Finally, SVM yields about 98% of accuracy with the liner prediction rate of 0.700 for classify different personalities. In future, the dataset is increased by adding more people and different algorithm is used for face verification systems.

ACKNOWLEDGEMENT

The authors can acknowledge any person/authorities in this section. This is not mandatory.

REFERENCES

[1] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3d morphable model. FGR, 2002.

[2] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few too many: Illumination cone models for face recognition under variable lighting and pose. PAMI, 23(6):643–660, 2001.

[3] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition? InWorkshop on Empirical Evaluation Methods in Computer Vision, December 2001.

[4] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, and W. Worek. Preliminary face recognition grand challenge results. FGR 2006, pages 15–24.

[5] G. Huang, M. Jones, and E. Learned-Miller. LFW results using a combined Nowak plus MERL recognizer. 2008.

[6] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. UMass Amherst Technical Report 07-49, October 2007.

[7] E. Nowak and F. Jurie. Learning visual similarity measures for comparing never seen objects. CVPR, 2007.

[8] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. UMass Amherst Technical Report 07-49, October 2007.

[9] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In Real-Life Images workshop at ECCV, 2008.

[10] G. W. Cottrell and J. Metcalfe. Empath: face, emotion, and gender recognition using holons. InNIPS, pages 564–571, 1990.

[11] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. SexNet: A neural network identifies sex from human faces. In NIPS, pages 572–577, 1990.

[12] B. Moghaddam and M.-H. Yang. Learning gender with support faces.PAMI, 24(5):707–711, May 2002.

[13] G. Shakhnarovich, P. Viola, and B. Moghaddam. A unified learning framework for real time face detection and classification. FGR, 2002.

[14] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. CVPRW, 05, 2003.

[15] J. Huang, X. Shao, and H. Wechsler. Face pose discrimination using support vector machines (SVM).ICPR, pages 154–156, 1998.

[16] A. Gallagher and T. Chen. Estimating age, gender, and identity using first name priors.CVPR, 2008.

[17] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images", IEEE Trans. on PAMI, VOL. 21, NO. 12, 1999, pp.1357-1362.

[18] C. Padgett and G.W. Cottrell, "Representing Face Images for Emotion Classification", in Proc. Conf. Advances in Neural Information Processing Systems, Denver, USA, 1996, pp. 894-900.

[19] T. Otsuka, and J. Ohya, "Spotting Segments Displaying Facial Expressions from Image Sequences Using HMM", in IEEE Proc. Int'l Conf. on FG , Nara, Japan, 1998, pp. 442-447.

[20] M. S. Bartlett, G. Littlewort, I. Fasel, et al., "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction", in IEEE CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction , Wisconsin, USA, 2003.

[21] R. Brunelli and T. Poggio. Hyperbf networks for gender classification. In Proceedings of the DARPA Image Understanding Workshop, pages 311–314,1992.

[22] B. A. Golomb, D.T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In Advances in Neural Information Processing Systems, pages572–577,1991.

[23] Garrison W. Cottrell. Empath: Face, emotion, and gender recognition using holons. In Advances in Neural Information Processing Systems, pages564–571,1991.

[24] S. Tamura, H. Kawai, and H. Mitsumoto. Male/female identification from 8×6 very low resolution face images by neural network. Pattern Recognition, 29(2):331–335,1996.

[25] Laurenz Wiskott, Jean-MarcFellous, Norbert Kruger, and Christoph Von der Malsburg. Face recognition and gender determination. In Proceedings of the International Workshop on Automatic Face and Gesture Recognition, pages92–97,1995.

[26] S. Gutta, H. Wechsler, and P. J.Phillips. Gender and ethnic classification. In Proceedings of the IEEE International Automatic Face And Gesture Recognition, pages194–199,1998.

[27] B.Moghaddam and Ming-Hsuan Yang. Gender classification with support vector machines. In Proc. Of Int'l Conf. on Automatic Face and Gesture Recognition, pages306–311, Grenoble, France, March2000.

[28] Cortes, C. and Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20 (3): 273–297. 1995.

[29] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In Proceedings of the Conference on Learning Theory, COLT'92, pages 144-152, 1992.

[30] Ana Merchan and Eddy Mayoraz, "Combination of binary classifi- ers for multi-class classification". IDIAP-Com 02, IDIAP, 1998.

[31] Miguel Moreira and Eddy Mayoraz. Improved pairwise coupling classification with correcting classifiers. IDIAP-RR 9, IDIAP, 1997.

[32] E. J. Bredensteiner and K. P. Bennett. Multicategory classification by support vector machines. Computa-tional Optimizations and Applications, pages 53–79, 1999.

[33] E. Mayoraz and E. Alpaydin. Support vector machines for multi-class classification. In IWANN (2), pages 833–842, 1999

[34] J. Weston and C. Watkins. Multi-class support vector machines. In M. Verleysen, editor, Proceedings of ESANN99, pages 219–224, Brussels, 1999.