A Hybrid Approach for Gender Classification of Web Images

Muhammad Usman Khan
Electrical Engineering Deptt
K.P.K University of Engineering
& Technology
Kohat Campus
Kohat, Pakistan

Hafiz Adnan Habib, PhD.
Computer Engineering Deptt
University of Engineering &
Technology
Taxila, Pakistan

Nasir Saleem
Institute of Engineering &
Technology
Gomal University
Dera Ismail Khan, Pakistan

ABSTRACT

In recent times, gender recognition of facial images has achieved lots of attraction. It can be useful in many places e.g. security, web searching, human computer interaction etc. In this paper, an approach containing both face detection and gender classification tasks has been proposed. In face detection part, Haar features have been chosen to present appearance features along with Ada-Boost technique to target strong and powerful features in cascaded form. For gender classification, Bayesian Classifier has been used where image is analyzed in blocks/patches form. The blocking technique is same as used in DCT approach. Experimental results have shown that proposed approach is effective and robust with changes in pose (some degree), expressions and illumination.

Keywords: Haar Features, Ada-Boost, Bayesian Classifier, DCT (Discrete Cosine Transform).

1. INTRODUCTION

If we see, then gender classification is a two ways classification issue, where one will tell about an image whether it is of a male or female. It is a very light task for human beings, but complex for machines like computers and robots. In order to reply for people's growing demands for authentic, reliable, easy to use and individualized services, the machine vision area has provided approaches like recognition of a face, recognition of gestures, and gender recognition that can play an important role in near future [1].

In start, gender recognition problem was entirely based on cognition and psychology [2, 3]. But in modern and scientific age, people have started thinking about this problem more technically.

If we consider the type of features that can be used for gender recognition purpose, then past knowledge can have two types: geometric (local) and appearance (global) based features. There are lots of techniques/algorithms that have been used in near past based on these features.

Scale Invariant Feature Transform (SIFT) is a mathematical algorithm for feature extraction and proposed by David Lowe and it can deal with misalignment and affine transformation [4]. SIFT is a mathematical approach to extract those points which are in our desired features from images.

In [5], block-based DCT had been used for extraction of features. In the DCT-based method, the top left square subset of DCT coefficients are used for classification purpose. The coefficients obtained after DCT implementation can be used latterly for classification purpose using any strong/weak/cascaded classifier.

The Principle Component Analysis (PCA) is another technique used for feature extraction and is also known as the K-L transform mostly used by researchers for signal and image processing [6], [7]. PCA can mostly be used in supervised and unsupervised learning.

Representative works in the field of gender classification include shunting inhibitory convolutional neural network with an accuracy rate of 85.7% [8] back propagation neural network along with PCA got an accuracy of more than 90% [9]. A new version of Support Vector Machines (SVM), called sparse SVM achieved very good result of 94.42 % with Japanese face images [10]. Automatic Ada-Boost system can get better accuracy rate than SVMs [11]. Other work includes LUT-based Ada-Boost method [12], KNN algorithm [13], Additive Logistic model [14] and Principle Geodesic Analysis approach [15] with comparative performances.

2. METHOD

Our proposed algorithm has two core modules: the first one detects a face in a given image and will crop and save it as a new image. If it doesn't find any face, it will report an error. The second one will use this detected face image as an input image; the final output of this module will show the conclusion.

2.1 Face Detection
2.1.1 Ada-Boost: The Boosting Algorithm

Ada-Boost short for Adaptive Boosting is a very well known machine learning algorithm, designed by Yoav Freund and Robert Schapire. It is a type of algorithm which is mostly used in conjunction with other learning algorithm(s), so that the performance of that algorithm(s) can be improved. In our technique, we have used Ada-Boost algorithm in combination with Haar type features which are also weak classifiers.

Ada-Boost technique takes as input a training set

$S = \{x_1y_1, x_2y_2\ldots x_ny_n\}$, where each x_i belongs to a domain X, and similarly each label y_i belongs to Y. Here in our work, we are only focusing on the binary case where $Y = \{0, 1\}$.

The main idea behind boosting is to use the weak learning of features which has been calculated, in order to form a highly correct prediction rules by calling the weak learner repeatedly which has been processed on different training examples.

2.1.2 Haar Cascaded Classifier

The Haar classifier is also a machine learning algorithm for the detection of visual objects, originally given by Viola & Jones. This technique was originally designed to detect faces but it is not only limited to detect faces. It can also detect other objects as well in an image. The biggest advantage of Haar classifier is that it quickly skips those regions which are not a part of object of interest.

The main part of Haar cascade classifier for detecting the object is Haar like features. These features don't use pixel intensity values but use contrast difference between adjacent group of pixels which will be rectangular or square [16]. Some Haar like features are shown in the figure 1, 2, 3. The value of a Haar feature can be calculated by difference between sums of pixel within black and white rectangular regions, i.e.

$F(x) = Sum_{black\ rectangle}$ (pixel gray level) $-\ Sum_{white\ rectangle}$ (pixel gray level)

To compute rapidly Haar like feature value, the concept of integral image is used. As shown in figure 4, integral image at location (x, y) contains sum of pixel values above and left of (x, y) inclusive:
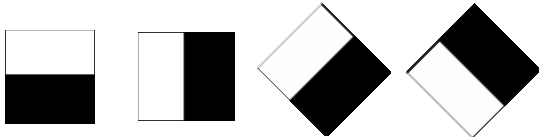
$$P(x,y) = \sum_{x' \le x, y' \le y} i(x', y')$$
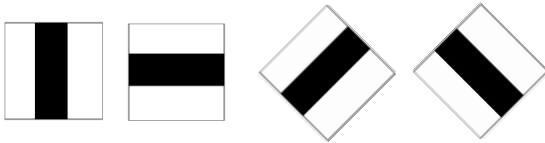
(1)

Figure 1: Two Rectangle Haar Features

Figure 2: Three Rectangle Haar Features
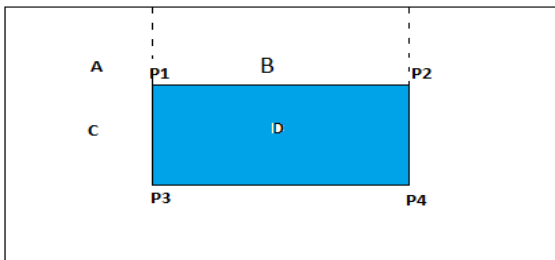

Figure 3: Four Rectangle Haar Features

Figure 4: Calculation of Integral Image

The sum of pixel values within "D":

$P_1 = A, P_2 = A+B, P_3 = A+C, P_4 = A+B+C+D$

$P_1+P_4-P_2-P_3 = A+A+B+C+D-A-B-A-C = D$

By using Haar like features, the face detection can be performed by a cascade as shown in figure 5. In that cascade, an image will be classified as a human face if it passes all the conditions i.e. $\{f_1, f_2…, f_n\}$. If at any stage, any of the condition becomes false then the image will not contain a human face and will not pass to further stages.

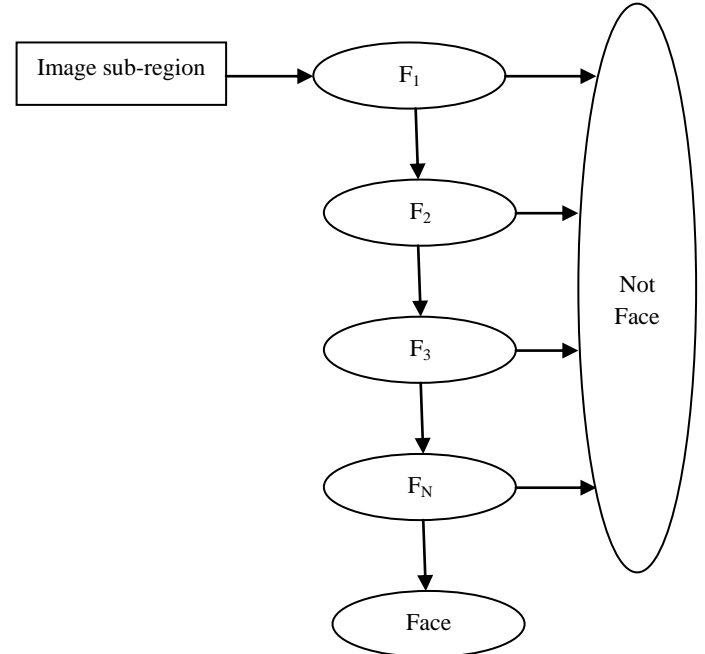**Figure 5: The Cascade Classifier for Face and Non-Face**

2.2 Gender Classification
2.2.1 Block Building
For gender recognition purpose, we have used a technique which is a well known feature of DCT. In DCT, the whole test image is broken up into a regular grid of blocks which will not overlap each other. Every block is handled separately and gives information which will be independent about class label (Male, Female). In our approach, at start, we have a library which is predefined instances of object. This library can be considered as a palette from which image blocks are obtained. A link between blocks of test image and blocks in library will be explored to find out class of test image.

2.2.2 Bayesian Classifier
2.2.2.1 Inference
In inference, we will assign a class label (e.g. C) to the input test image and possibly we can have K classes i.e C ∈ {1, 2 … K}. The overall inference procedure can be explained in phases as shown in figure 6.
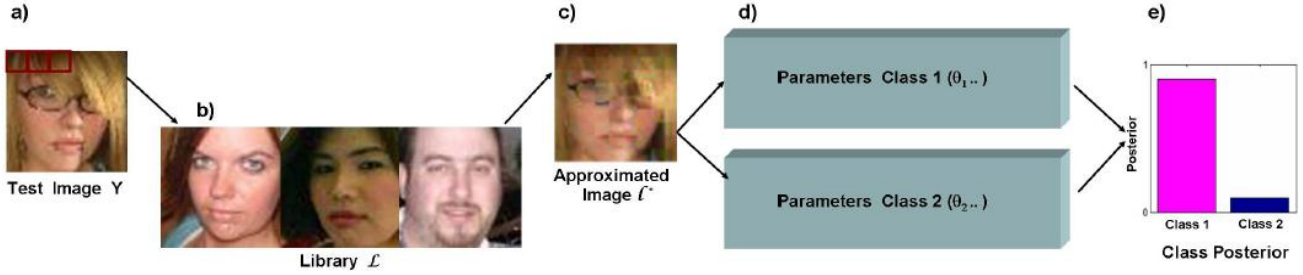
Figure 6: Inference Procedure

In figure 6, we have five phases and these phases will define the whole inference procedure. Phase (a) is the input image which will be broken down into a regular grid of blocks. Let call this input image 'Y' where Y = [y_1, y_2....y_p]. Phase (b) contains a large library 'L' consists of face images with class label 'C' where C $\in$ {1, 2 … K} which will be used to approximate the class of given input block. Each block from library 'L' will have a class. Here each image in library will have same number of blocks as in input image. Phase (c) will again have input image but now with its approximated classes for each of its blocks. Phase (d) contains a parameter Ө which will help to gather all those blocks of phase (c) which belong to same class. Phase (e) contains a Bayesian classifier which will calculate posterior probability of each class to determine/approximate the overall class of input image 'Y'.

The above proposed model can be trained latterly from 'I' input image examples where each image has a class X_c in 'K' classes and block from input image will be called p_{th} and from library will be called i_{th}. The library overall will contain 'L_l' blocks where l $\in$ {1…N} where N represent the index number of block of library. The result will be posterior probability over class label C. We compute it by Bayes' rule

$$P_r(C = c|Y, X_\bullet) = \frac{\prod_{p-1}^{P} P_r(y_p|C = c, x_{\bullet cp}) P_r(C = c)}{P_r(Y)}$$

(2)

Here we will assume that the test block y_p will be independent. The symbol • shows all values of an index, so $X_\bullet$ = {X_1…X_K} represents training images from all K classes and $x_{\bullet cp}$ presents p_{th} block from all training images I from c_{th} class. Though likelihood is dependent on library, it doesn't depend on parameters of model Ө. Here we apply Bayesian technique and marginalize the result on parameters of model, so likelihood terms can have a form like:

$$P_r(y_p|C = c, x_{\bullet cp}) = \int P_r(y_p|\Theta_{cp\bullet}) P_r(\Theta_{cp\bullet}|x_{\bullet cp}) d\Theta_{cp}$$

(3)

Here $\Theta_{cp\bullet}$ means all parameters linked with p_{th} patch of c_{th} class.

In order to compute required likelihood, we will find index l^* of library first which will mostly resembles test block y_p. Here we will assume about test block that it is a Gaussian corruption of library block and with the help of maximum posteriori estimation, we can find l^*

$$l^* = \arg max_l \mathcal{G}_{yp}[\mathcal{L}_l; \sigma^2 I]$$

(4)

The likelihood can be found as

$$P_r(y_p|\Theta_{cp\bullet}) = P_r(l^*|\Theta_{cp\bullet}) = \Theta_{cpl^*}$$

(5)

2.2.2.2 Training

Here we will show how to make use of training data $x_{\bullet cp}$ from p_{th} block of every image that belongs to c_{th} class in order to get posterior distribution $P_r(\Theta_{cp\bullet}|x_{\bullet cp})$ over relevant parameters $\Theta_{cp\bullet}$. The posterior distribution was seen in equation (3). To compute posterior distribution over parameter $\Theta_{cp\bullet}$, we have a second application of Bayes' rule which is:

$$\Pr(\theta_{cp\bullet}|x_{\bullet cp}) = \frac{\Pr(x_{\bullet cp}|\theta_{cp\bullet}) \Pr(\theta_{cp\bullet})}{\Pr(x_{\bullet cp})}$$

(6)

In order to simplify the process, we will just focus on single block from P regular blocks and one class from K classes and will remove indices p and c. The equation (6) will be like:

$$\Pr(\theta_\bullet|x_\bullet) = \frac{\Pr(x_\bullet|\theta_\bullet) \Pr(\theta_\bullet)}{\Pr(x_\bullet)}$$

(7)

Where $x_\bullet$ = x_1…x_n is all training data for this block and this class and $\Theta_\bullet$ = Θ_1…Θ_N is a vector of N parameters linked with every location in library for this block and class.

Now if we want to compute likelihood of i_{th} training example, called x_i with parameter $\Theta_\bullet$ we will need to search out nearest similar library block $\widehat{l_i}$ where

$$\widehat{l_i} = \arg max_l \mathcal{G}_{x_i}[\mathcal{L}_l; \sigma^2 I]$$

(8)

Now if we consider equation (5), then we have

$$P_r(x_i|\Theta_\bullet) = P_r(\widehat{l_i}|\Theta_\bullet) = \Theta_{\widehat{l_i}}$$

(9)

If we take whole training data $x_\bullet$, then equation (9) will become

$$P_r(x_i|\Theta_\bullet) = \prod_{i=1}^{I} \Pr(x_i|\theta_\bullet) = \prod_{i=1}^{I} \Theta_{\widehat{l_i}} = \prod_{l=1}^{N} (\theta_l)^{fl}$$

(10)

Where fl is defined as

$$fl = \sum_{i=1}^{I} \delta_{\hat{l}_i = l}$$

(11)

Where $\delta_{\hat{l}_i = l}$ will return one if subscripted expression $\hat{l}_i = l$ is right, otherwise zero. So fl will define the total number of times, the nearest resembling block came from library in training process.

Now we will define prior over parameter Θ in equation (7). For that we have chosen a Dirichlet prior of the form

$$\Pr(\theta_\bullet) = \frac{\Gamma \sum_l \alpha l}{\prod_l \Gamma(\alpha l)} \prod_{i=1}^{N} (\theta_l)^{\alpha l - 1}$$

(12)

Where Γ is used to denote Gamma distribution and $\{\alpha_1 \ldots \alpha_N\}$ are parameters of this Dirichlet distribution.

Now by putting the likelihood calculated in equation (10) and the conjugate prior term of equation (12) in Bayes' rule defined in equation (7), we will get a relation for posterior distribution over parameters in the form of a Dirichlet distribution:

$$\Pr(\theta_\bullet | x_\bullet) = \frac{\Gamma \sum_l (\alpha l + fl)}{\prod_l \Gamma(\alpha l + fl)} \prod_{i=1}^{N} (\theta_l)^{fl + \alpha l - 1}$$

(13)

We calculate these distributions for every P block in regular grid and for every class in K classes.

2.2.2.3 Calculation of Likelihood Integral

In last, we put posterior distribution over parameters $\Pr(\theta_{cp\bullet} | x_{\bullet cp})$ by resuming indices p and c and integrate over $\Theta_{cp\bullet}$ which will give us:

$$\Pr(y_p | C = c, x_{\bullet cp}) = \frac{f_{cpl^*} + \alpha_{l^*}}{\sum_l (f_{cpl} + \alpha_l)}$$

(14)

3. EXPERIMENTAL SETTINGS

In our experiment, gender of faces has been checked automatically by proposed technique. A database of images has been made by populating websites which contains images of both males and females. These images are from uncontrolled environments and show many changes in illumination, pose, expression, pose, scale and background. Faces were detected and cropped using above mentioned method i.e. Ada-Boost technique.

In our proposed approach, the test dataset have 200 face images, equally divided in each gender and library contains 240 face images with same rule. We got 95% accuracy rate for male faces and 91% for female faces. Figure 7 shows examples of accurately classified frontal faces of male and female. Figure 8 shows correctly classified images of male and female with variations in pose. Figure 9 shows those examples which are misclassified. Figure 10 shows a sample of database images which only contain face portion.

4. PERFORMANCE COMPARISON

Figure 11 and 12 contain graphs of classifications for male and female images respectively. In each graph, vertical axis shows accuracy in terms of percentage while horizontal axis shows total number of images used. The red changing line in each graph shows variation in accuracy of algorithm. The failure of algorithm is shown by down peaks in each graph. Each down peak is representing number of image. As said earlier, more accuracy was found in case of male images, as compared to female images. Therefore there are more down peaks in female classification graph.

For verification purpose, a comparison between performance of our proposed algorithm and performance of well known approach like support vector machines (SVMs) and linear discriminant analysis has been done. Literature and experiments show that SVMs are not suitable for large databases and it will be very difficult to train it with high resolution images e.g. 60x60 or above. The reasons are heavy processing, time consumption, memory requirements etc. For verification purposes, we have used both (i) the training images at high resolution and (ii) the training images at low resolution. It is like images used in [17].

For high resolution images, linear SVMs and non-linear SVMs with an RBF kernel have an accuracy rate of 78.8% and 77.8% respectively. When we have tested our algorithm performance with linear and non linear SVM, our proposed method has considerably better performance. For low resolution images, linear and non-linear SVMS has 78.7% and 82.4% accuracy rate respectively which is still low. For same images cases, linear discriminant analysis have got 78% accuracy rate [19], which is also low. If we see, it is clear that above mentioned algorithms didn't get an accuracy rate above 89% but our proposed algorithm did that.

Figure 7: Correctly Classified Faces

Figure 8: Correctly Classified Faces

Figure 9: Misclassified Faces

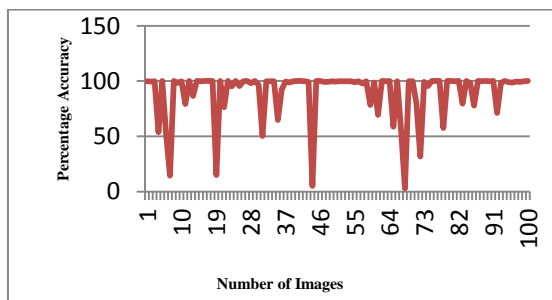**Figure 10: Sample Data Set of Cropped Faces**

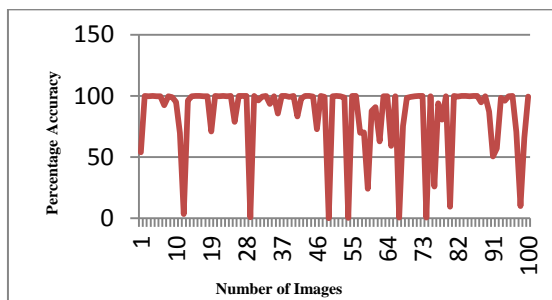**Figure 11: Graph for Male Classification**

Figure 12: Graph for Female Classification

5. CONCLUSION AND FUTURE WORK

In this paper, a new hybrid approach by fusing face detection and gender classification was presented. We use Ada-Boost algorithm for face detection and cropping purpose combined with Haar features. The cropped faces are latterly classified by Bayesian classifier. Experimental results demonstrate that the hybrid method obtains a higher accuracy. Moreover, the novel classification method shows effectiveness and robustness regarding expression, illumination and pose variation in some degree.

There are still some aspects that deserve further study. First of all, the face detection part needs to detect and crop faces with more variation in pose, illumination and expression. Also there is a need to detect multiple faces in parallel. The gender classification part needs more tuning for better classification in cases of aged and teenage persons. The time consumed for face detection and gender classification can be decreased if a further optimization is made in order to support real time applications.

6. REFERENCES

[1] Ziyi Xu, Li Lu and Pengfei Shi, "A Hybrid Approach to Gender Classification from Face Images", Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China, IEEE, 2008.

[2] A. Golomb, D. T. Lawrence, and T. j. Sejnowski, "SEXNET: A neural network identifies sex from human faces", Neural Information Processing Systems, 1991, pp: 572-577.

[3] G. w. Cottrell and J. Metcalfe. EMPATH: "Face, emotion, and gender recognition using holons". Neural Information Processing Systems, 1991, pp: 564-571.

[4] Lowe David G., "Object Recognition from Local Scale-Invariant Features", International Conference on Computer Vision, Canada, 1999.

[5] Podilchuk and X. Zhang, "Face Recognition using DCT Based Feature Vectors", In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 4, Pp. 2144-2147, 1996.

[6] V. Algazi, K. Brown, and M. Ready, "Transform representation of the spectra of acoustic speech segments with appliances, part I: General approach and application to speech recognition," IEEE Trans. Speech Audio Process., vol. 1, pp. 180–195, 1993.

[7] L. Sirovich and L. Keefe, "Low dimensional procedure for characterization of human faces," J. Opt. Soc. Amer., vol. 4, pp. 519–524, 1987.

[8] F. H. Tivive and A. Bouzerdoum, "A Gender Recognition System Using Shunting Inhibitory Convolutional Neural Networks", Faculty of Informatics, University of Wollongong, 2006.

[9] P.Latha, Dr.L.Ganesan, Dr.S.Annadurai, "Face Recognition using Neural Networks", Signal Processing: An International Journal (SPIJ), Volume 3, 153-160, 2009.

[10] Costen, N., Brown, M., Akamatsu, S., "Sparse models for gender classification". Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'04), 2004, pp: 201–206.

[11] Shakhnarovich, G., Viola, P.A., Moghaddam, B., "A unified learning framework for real time face detection and classification", Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'02). IEEE, 2002, pp: 14–21.

[12] B. Wu, H. Ai, C. Huang, "Real Time Gender Classification", Multi-spectral Image Processing and Pattern Recognition, 2003.

[13] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), p. 511, Volume 1, 2001.

[14] Simon J.D. Prince and Jania Aghajanian, "Gender Classification in Uncontrolled Settings Using Additive Logistic Models", Department of Computer Science, University College London.

[15] J.Wu,W. Smith, and E. Hancock, "Gender classification using shape from shading", In BMVC 2007, pages 499–508, 2007.

[16] Zhiming Liu, Jian Yang, Chengjun Liu, "Extracting Multiple Features in the CID Color Space for Face Recognition", Dept. of Computer Science, New Jersey Institute of Technology, USA, IEEE Explore, April 2010, pages 2502 – 2509.

[17] B. Moghaddam and M. Yang, "Learning Gender with Support Faces", PAMI, pp. 707–711, 2002.

[18] A.A. Efros andW. T. Freeman, "Image quilting for texture synthesis and transfer", Proc. SIGGRAPH, pp. 341-346, 2000.

[19] P. Felzenszwalb, D. McAllester and D. Ramanan, "A discriminatively trained, multi-scale, deformable part model", CVPR, pp. 1-8, 2008.