Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 10, pp.59193-59198, October, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

AUTOMATIC FACE IMAGE ANNOTATION USING MACHINE LEARNING TECHNIQUES

***Nagila, S. and Myna, A. N.**

Ramaiah Institute of Technology, Department of Information Science and Engineering, Bangalore, India

ARTICLE INFO

Article History:
Received 28th July, 2017
Received in revised form
18th August, 2017
Accepted 22nd September, 2017
Published online 31st October, 2017

Key words:

Face image annotation, Viola and Jones, Local Binary Pattern, Support Vector Machine.

ABSTRACT

Face image annotation refers to automatically assign labels to objects in a face image. Assigned labels may be gender, age, name, etc. For annotating the human faces, wide variety of algorithms and techniques are used in computer vision. In this paper, a real time face image annotation system which involves face detection, gender and age identification is proposed. For detection of human faces in an image Viola and Jones Object detection technique is used. For the detected human faces, features are extracted using the process of Local Binary Pattern. For classification of human faces based on their gender and age, Support Vector Machine is used by considering extracted features as support vectors. The main application of this work is to count the number of human faces who have viewed a particular advertisement. The face database consists of sample face dataset of varying age group and gender. The experimental results show better accuracy for gender and age classification.

Citation: Nagila, S. and Myna, A. N. 2017. "Automatic face image annotation using machine learning techniques", *International Journal of Current Research*, 9, (10), 59193-59198.

INTRODUCTION

The main goal of an automated image annotation model is assigning of pre-loaded tags for test images. Automatically analyzing, and assigning meaningful annotations for the human faces is a challenging task. The proposed work captures the human faces who are viewing a particular video frame and finds the number of people who viewed the video which is displayed on screen. Manually identifying a person's age, gender is a time consuming task. Instead, automatic face annotation model is built which automatically annotates person's gender and age. Hence there is an improvement in accuracy when compared with manual approach. Face Annotation involves Image Capturing, Face Detection, Facial Feature Extraction, Classification, Face Annotation and Generating Count (Priyanka *et al*., 2016).

***A. Capturing the face image*:** As the person stands in front of the video, the web cam automatically identifies the location of the person and captures the image. The system captures real-time images and further analysis of images and videos is carried out in real time.

B. Face Detection: Detecting a face image in a sequence of images/videos and locating a face area within the image is a challenging task as there will be a wide variety of changes in the face image. These changes, while capturing the face image could be due to variation in light conditions or posture. There are a wide variety of algorithms used to detect the human face in an image. Meir *et al*, 2003, used AdaBoost algorithm for detection of human faces. To generate single composite learner AdaBoost uses multiple iterations. In this approach both positive and negative examples are tested. The drawback of this approach is that in each iteration, the training data yields slow output. Henry *et al*, 1998 developed a neural network based frontal face detection model. In their work, an image is considered and a small sliding window is attached to the image. Using neural networks those windows are examined and checked for the presence of a face inside a particular sliding window. This model arbitrates between many numbers of neural network, which directly improves the performance of detecting the face. For detecting the moving objects over the background. Ahonen *et al*, 2006 used Local Binary Pattern, facial features are extracted from local regions where the model identifies the key points in the targeted region. Their proposed system was not so sensitive to small changes in locating face images in an image and comparison is carried out between local binary pattern and local texture description. Viola *et al*, 2003 worked on real-time face detection. The computational time is minimized when detecting faces in an image which in turn achieves high detection accuracy. The method concentrates on face detection and not on the identification or recognition. This is useful when there is more importance given to the number of frames analyzed per second. This method gives high accuracy while detecting the face in an image. It generates very low false positive rate (Malardalens Hogskola, 2015). The whole process of face

Corresponding author:* **Nagila, S.
Ramaiah Institute of Technology, Department of Information Science and Engineering, Bangalore, India.

detection is undertaken in real-time. Among these techniques, Viola and Jones algorithm gives best results. If the quality of the image is low, then the image needs to be pre-processed. To satisfy the requirements of feature extraction, first the image has to be pre-processed. (Krishna *et al*., 2014) analyzed the effect of pre-processing. This step is carried out prior to feature extraction process. For the pre-processing of an facial image (Hemavathi *et al*., 2016) worked on Median and Gabor filters. Basically pre-Processing involves- Color Normalization, Noise Reduction, Edge Detection and Histogram Equalization. To reduce the noise and to normalize the color in the image, Median filters are used. To enhance the edge of an image Gabor Filter is used. With the help of these Hybrid filters, the quality of an image is enhanced and the pre-processing efficiency is increased. To enhance the face image, a well known image enhancement technique used is Histogram Equalization. This process increases the contrast that is present on a face image. Graphical representation of color distribution in an image is explained by (Hemavathi *et al*., 2016) where the graphical image exhibits variety of distribution of data. After the equalization of an image there will be improvement in the quality of the image.

C. Facial Feature Extraction: After the pre-processing of face image, features are extracted through Neural Networks, Fuzzy Extraction Methods, Local Binary Patterns, Knowledge based approach or Normalization. To extract the features present in an image (Brijesh Patel *et al*., 2014) developed a prediction model. In that model, features are extracted using Active Appearance model. One of the robust methods for extraction of features is geometry based approach, where the features are extracted using geometric information of face component. In statistical knowledge and generic visual approach features are extracted. The extraction of features is done based on generic knowledge of facial components. This shows that with the help of Knowledge Based Approach, features of face can be extracted. A Multi-View model is developed by the authors (Hui-Cheng *et al*., 2007) and they used Local Binary Pattern for extracting the features present on the face. The proposed model concentrated on shape and texture information of a facial image. To extract the patterns in an image Local Binary Pattern is used where the image is resized to small regions and histogram are extracted. These Histograms are concatenated to a single vector that represents the facial image. Among these approaches, Local Binary Pattern gives accurate results while extracting the features.

D. Classification: After the face is detected, it has to be classified based on gender and age. There are various methods used to classify the human faces. An experiment was carried out, (Malkarthekar *et al*., 2009) where the experimental analysis for classification of facial images was done. The experiment was performed on different set of facial images with different expressions. Fisher Discriminate and Euclidian Distance for finding out the closet match is used. A survey on age group prediction was carried out. (Brijesh Patel *et al*., 2014) They reported that under controlled environment conditions age can be accurately estimated but it will not achieve same level of accuracy in real-time examples. The research was carried out on semi-supervised approach which predicts the approximate age of a person. A fast and efficient gender and age estimation model was developed. (Hlaing Hlake Khaung TIN, 2011) The research was carried out in 3 different steps: First the shape of the face image is described. Next the features are extracted in feature extraction phase, where the geometric features are being evaluated by keeping the ratio of distance between the eye, nose, mouth. Finally based on Principal Component Analysis method and Geometric feature based methods classification of face image based on gender and age is carried out. Another approach to classify the face images based on gender and age is Support Vector Machine Learning Algorithm (Michaela Romanca, Dr. Rajeev Srivastava).

E. Annotation: It involves labeling the face images with their relevant specifications (gender and age). Assigning labels to the face images improves the efficiency of detecting face images.

F. Generate Count: In involves counting the number of people who view the video for a few seconds.

Face Detection

Viola and Jones object detection framework provides maximum competitive face detection rates in real time. To solve the problem of face detection, the machine must be trained to detect variety of face objects in an image. The machine is trained in such a way that it identifies the location of face region in an image. Viola and Jones algorithm uses 3 steps for processing of face image:

Haar-like-feature

Haar-like-features are utilized in Viola and Jones algorithm, that are used to detect the features present in face image which are formed by the scalar product of image. According to the author (Papageorgin *et al*., 1998), the simple basic functions where used for features: Haar-like-features. Specifically, 3 different types of features are used: (i) two-rectangle features- sum of the pixels is calculated, between the two different rectangular regions (vertically and horizontally), (ii) three-rectangular features- sum of centre rectangle is subtracted from outside two rectangles (iii) four-rectangular features- difference is calculated between the diagonal pairs of rectangle as shown in the Fig 1.

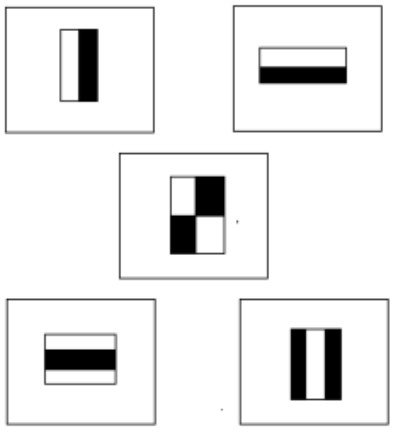

Fig. 1. Example of rectangular features shown relative to encounter detection window

The detection framework is universally employed with features that involve the sums of image pixels within rectangular area.

Calculating the integral image

To create a integral image, every pixel is evaluated which is the sum of all the pixels above and the left of the pixel. Integral image is defined as:

$$ii(x,y) = \sum_{x^{|} \leq x, y^{|} \leq y} i(x^{|}, y^{|})$$

where ii(x,y) is the intensity of the gray scale image at pixel (x, y). Using the integral image as illustrated in Fig. 2, the sum of the intensity pixels of any rectangular area ABCD can be calculated with only four array references.(Monali Chaudhari *et al*, 2015)

By using 4 values it is possible to compute the addition of entire pixels of any specific rectangle in an image.

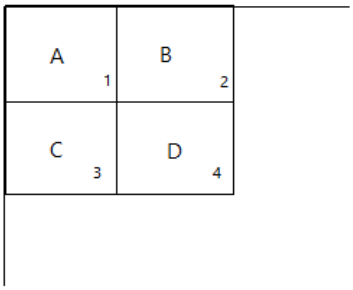

Fig. 2. Calculation of the integral image

Procedure to calculate the integral image:

Value of the integral image at location 1 is the sum of the pixels in the rectangle A
Value of the integral image at location 2 is the sum of the pixels in the rectangle A+B
Value of the integral image at location 3 is the sum of the pixels in the rectangle A+C
Value of the integral image at location 4 is the sum of the pixels in the rectangle A+B+C+D.

Intensity value for the rectangular region ABCD is calculated as follows:

$\sum_{x,y \in (A,B,C,D)} i(x,y) = ii(D) - ii(B+C) + i(A)$

Sum of Region = 4-(2+3)+1

Learning algorithm and Boosting

As stated previously, in Viola and Jones algorithm there will be approximately 160000+ feature values within detector at 24*24 window base resolution which need to be calculated. Haar-features check in all possible combinations of positions, size and degree to confirm whether all the features are relevant. But it is understood that only few set of features will be useful among all the features to identify a face in an image. there are presence of many irrelevant features present in an image so to eliminate it ADA Boost algorithm is used. ADA Boost identifies certain number of features from 160000+ features (TSANG TAT MAN). Later weights will be assigned. Features along with weights are evaluated. Later linear combination of all the output is verified and checks whether that value exceeds certain threshold where threshold is a weak classifier on a single Haar-like feature. Linear combination of features is going to identify the presence of face in an image and formulated using:

$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots$

F(x) – strong classifier

$\alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x)$ – weak classifier

$\alpha_1, \alpha_2, \alpha_3 \ldots$ assigned weights, $f_1, f_2, f_3 \ldots$ assigned features

In the whole process boosting is considered to be the most time consuming procedure. We apply relevant weights for the features, the process is called weak classifier and by combining the weights and features we get a strong classifier. The output of weak classifier is 1 which means performance is good and identifies the features when applied on image, 0 is assigned when the feature is not identified in an image.

Attentional Cascade

Cascade classifier is formed with stages that consists of strong classifier. Due to the calculation 160000+ features at a time which will decrease the computational time and cost. Instead cascade classifier is used which readily decreases the computational time and performance and the newly constructed boosting classifier neglects many of negative sub-windows at the time of detecting the positive instances.

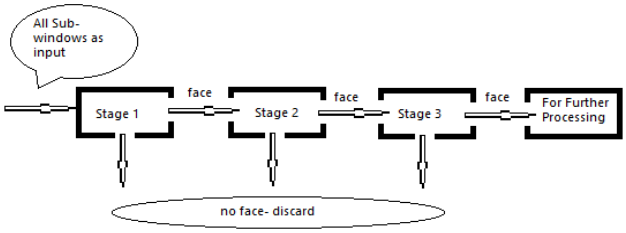

Fig. 3. Cascade classifier with stages

Cascade classifier (Hiyam Hatem *et al*., 2015) is built with stages that distributes the features equally to the stages which increase the complexity. It is applied on 24*24 window of a given image, then need to check whether face features are identified or found in the sub-window. If the sub-window classifies the image as face not found, then the sub-window is discarded. And the sub-window moves to the next location of 24*24 pixel window. Consequently, if the sub-window classifies as face found then the classifier will pass the face for the next stage for further computation. Viola and Jones algorithm precisely accepts many number of false positive rates in initial stages. Therefore, false negative rates are slightly less when the final stage of cascade classifier is arrived.

Local Binary Pattern

Various feature extraction methods are evolved for extracting features from a face image and one among them is Local Binary Pattern. LBP has been used for age and gender classification based on texture. Face image is divided into several small regions through which extraction of features is carried out. Each features consists of binary patterns. With the help of binary patterns, the surrounding pixels information is calculated. Each region consists of histogram based on features

and are concentrated into a single big feature histogram, where in it represents the information of whole face image. Consider a face image, divide the face image into 8*8 pixel format and each block is represented with 3*3 pixel matrix block which consists of 9 pixels at a time in a block and it is more concentrated at central pixel. LBP compares the neighboring pixel with central pixel. When the central pixel value is greater or equal to the central pixel then the value is replaced by 1 or if the central pixel value is lesser than the central value then the pixel value is replaced by 0, all the values are converted to binary value based on their central pixel which is considered as threshold. By using the binary pattern, decimal value is calculated. The whole procedure is based on following equation as given in (Ihsan Ullah *et al.*, 2014):

$$LBP_R^P(I_c) = \sum_{n=0}^{P-1} s(I_n - I_c)2^n$$

Where (I_c) and (I_n) are the values of center pixel and neighbouring pixel. And s value will be either 1 or 0. If 8-bit binary pattern is used, total of 2^8 ie 256 different binary patterns are formed. Totally 256 patterns are represented in different bins. Local Binary Pattern is implemented using uniform and non-uniform LBP. In case of uniform LBP, if binary pattern consists of two bitwise operators 0 & 1. There will be transitions from 1 to 0 or 0 to 1 and such transitions are called as 2 transition uniform LBP. If the transitions in binary bittern is more than two times then it is called as non-uniform LBP. Totally there are 58 uniform patterns represented in 58 bins (Andrey Rybintsev, 2017) and remaining binary patterns are non uniform patterns and they are represented in a single bin. So totally there are 59 patterns of LBP code formed and a histogram is created for all the 59 patterns in 59 different bins. The histogram generated helps in representing the each pixel is labeled which represent the information of face image.

Support Vector Machine

Support Vector Machine is used to classify the set of data. For a particular class, set of data will be distributed. Based on the characteristics of data, it need to be grouped and added to any of the class and classification of data is done based on the features. The data is spread over a vector space and is assumed that several classes are distributed in a vector space which is decision making approach. The class of data is separated using a hyper plane. The data points that lie on hyper plane are called as Support Vectors. (Suralkar *et al.*, 2012)

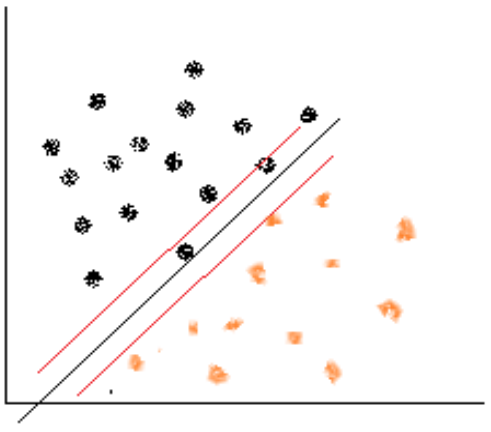

Fig. 4. Representation of support vectors in vector space

Proposed method

Step 1: Obtain images from live video stream
Step 2: Convert the RGB image into Gray scale
Step 3: Resize the input image
Step 4: Image is pre-processed using Histogram Equalization
Step 5: From image, face region is detected using Viola and Jones technique. The algorithm consists of four stages:

- Haar features selection
- Creating a integral image
- Ada Boosting training
- Cascading classifier

Step 6: Face region is detected and marked using a rectangular box
Step 7: Facial features are extracted using Local Binary Pattern algorithm

- Divide face detected region into N block based on N*N pixels
- Calculate LBP code for each block of N*N pixel
- Local LBP histogram for each block is built
- A single vector feature histogram is obtained by combining all the histograms of each block

Step 8: Interpolation is performed for each N*N block
Step 9: Local regions detected by LBP are edges, spots, corners, line tip are the texture features from which region of facial features are extracted
Step 10: Variety of textures are formed on face region and based on pre-defined features, extraction of features is performed:

- Features for age estimation -> wrinkle analysis and skin tone analysis
- Features for gender classification -> hair regions on face and skin tone analysis

Step 11: Classification of face images is performed based on gender and age using feature extraction and non-linear Support Vector Machine algorithm
Step 12: While classifying the face images, radial bias function is used as follows:

- To classify face images based on age rbf=0.6, while comparing the face images, if the comparison rate is more than rbf value then the face images is classified to that particular age
- To classify face images based on gender rbf=0.8, while comparing the face images, if the comparison rate is more than rbf value then the face images is classified to that particular gender.

Step 13: Verification of face images is performed.
Step 14: Alert messages are displayed
Step 15: Annotation of face images with specific features identified (age, gender) and displayed on rectangular box.
Step 16: Stop the whole procedure until the new image is captured.

RESULTS AND ANALYSIS

The implemented model helps in testing the performance of age classification and gender classification on different face images using Local Binary Pattern for feature extraction. For classification of age and gender a well known classifier is used – Support Vector Machine. To evaluate Support Vector Machine, a set of training set of images were trained and compared with test images. Accordingly the test images were annotated through which it boosts the system to generate the count of individual who has viewed the video screen. To note down the accuracy rate, the following calculation is done.

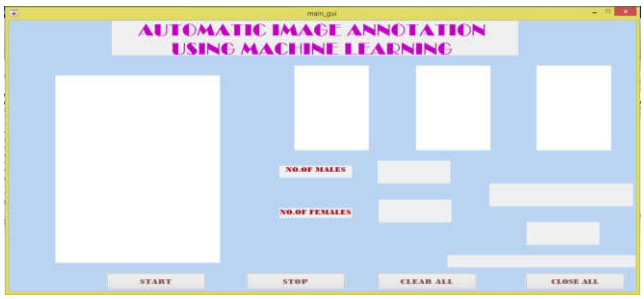

Fig. 5. The GUI of the proposed model

Fig. 6. The GUI with advertisement roll on

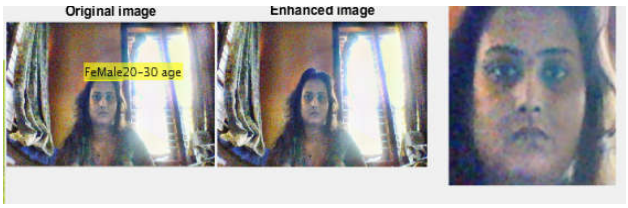

Fig. 7. The Captured image is later pre-processed and only face region is detected, cropped and resized

$$ACCURACY = \frac{\text{Number of people currently viewed video screen and identified}}{\text{Total number of people (test face profiles – male or female)}} * 100$$

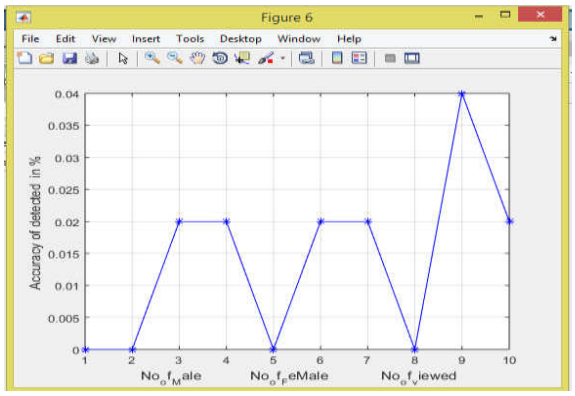

Fig. 8. Performance analysis of people viewing the video screen is plotted on a graph

People viewing the advertisement on video screen and generating the count:-

Table 1. The Accuracy Rate obtained for the procedure is 87.33%

	Male	Female	Total
ADD1	1	0	1
ADD2	0	2	2
ADD3	0	0	0
ADD4	2	1	3

Conclusion and Future work

A real time face detection and annotation system has been implemented which automatically detects the human faces using various conditions: like under various lighting conditions, frontal pose of the human, presence of glasses, facial hair and variety of hair style done by humans. The proposed method of face detection is found to be accurately detecting all kinds of single view frontal faces annotating and finally count is generated for the people who view the video screen. The whole work is divided into modules and functionality of each module describes the performance and accuracy rate which will be proportionally increased. Face detection is an interesting task and this application can be used for authentication for home security, for banking system where personnel identification is happening. Presently digital cameras are used to detect faces in auto focus. In future face detection and verification can be extended for student ID, driving license, aadhaar card. The main approach would be implemented for detecting and locating more than one face region in frontal view and it improves robustness and accuracy in tracking of human frontal view face detection. The technique SVM must be trained with large number of data sets which will give the output in more accurate way. Finally the approach can be extended other than frontal view face detection by using all other different angles of face detection approach can be implemented.

REFERENCES

Andrey Rybintsev, 2017. "Age Estimation from a Face image in a selected gender-race group based on ranking local binary patterns", Springer CrossMark, Complex Intell. Syst. 3: 93-104.

Brijesh Patel, Raghvendra 2014. "Gender recognition and Age-group Predication: A survey", *International Journal of Engineering and Computer Science*, Vol 03, Issue 12 December 2014: pp 9564-9567.

Constantine P. Papageorgiou, Michael Orem, Tomaso Poggio. 1998. "A General Framework for Object Detection", IEEE Conference Paper, Source: IEEE Xplore, pp 555-562.

Hemalatha, G. and C.P, Sumathi, 2016. "Preprocessing Techniques of Facial Image with Median and Gabor Filters", International Conference On Information Communication And Embedded System (ICICES2016), IEEE, 25-26 Feb. 2016

Henry A. Rowley, Shumeet Baluja and Takeo Kanade, 1998. "Neural Network-Based Face Detection" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol: 20, Issue: 1)

Hiyam Hatem, Zou Beiji and Raed Majeed. 2015. "A Survey of Base Features for Human Face Detection", *International Journal of Control and Automation*, Vol 8, No.5, pp 61-78.

Hlaing Htake Khaung TIN. 2011. "Gender and Age Estimation Based on Facial Images", ACTA TECHNICA NAPOCENSIS Electronics and Telecommunications, Vol 52, number 3, pp 37-40. www.researchgate.net/publication/263547472

HUI-CHENG LIAN and BAO-LIANG LU. 2007. "Multi-View Gender Classification using Multi-Resolution Local Binary Patterns and Support Vector Machines". *International Journal of Neural System*, Vol 17, No 06:pp 479-487.

Ihsan Ullah, Hatim Aboalsamh, Muhammad Hussain, Ghulam Muhammad, George Bebis, 2014. "Gender Classification from Facial Images using Texture Descriptors", *Journal of Internet Technology*, Vol 15, No.5 pp 801-811

Krishna Dharavath, G. Amarnath, Fazal A. Talukdar and Rabul H. Laskar, 2014. "Impact of Image Preprocessing on Face Recognition: A Comparative Analysis" IEEE, International Conference on Communication and Signal Processing, India, pp-631-63

M.D. Malkauthekar, Shubhangi Sapkal and Sangeeta Kakarwal, 2009. "Experimental Analysis of Classification of Facial Images", IEEE International Conference on Advanced Computing pp: 1093 – 1098

Malardalens Hogskola, 2015. A report on "Real Time Image Recognition of facial features for detecting a true breath". http://www.diva-portal.org/smash/get/diva2:822997/attachment01.pdf

Meir R. and G. R"atsch. 2003. "An introduction to boosting and leveraging" in Lecture Notes In Computer Science; pp 118-183; Machine learning by Springer , New York.

Michaela Romanca, a report on "Support Vector Machines in Face Detection Systems", "https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwjX_ZLR5OrWAhVKLY8KHQY6BdsQFgg1MAI&url=https%3A%2F%2Fpapers.nips.cc%2Fpaper%2F1609-support-vector-machines-applied-to-face-recognition.pdf&usg=AOvVaw1CyDiuFC0icUYiqnVIY7FC"

Monali Chaudhari, Shanta Sondur, Gauresh Vanjar, 2015. "A Review on Face Detection and Study of Viola and Jones method", *International Journal of Computer Trends and Technology*, Vol 25, pp 54-61.

Paul Viola and Michael J. Jones, 2003. "Robust Real-Time Face Detection", *International Journal of Computer Vision,* 57(2), pages 137-154.

Priyanka Chaudhari, Pallavi Nagpur, Bansi Patel, Shital Bhamre, 2016. "FAnn!!! Search Based Face Annotation" *International Research Journal of Engineering and Technology (IRJET)*, pages 904-908, Volume: 03 Issue: 05.

Rajeev Srivastava , experiment on "Pattern Classification using SVM Classifier in MATLAB". http://content.inflibnet.ac.in/data-server/eacharya-documents/548158e2e41301125fd790c3_INFIEP_58/47/ET/58-47-ET-V1-S1__svmclassifier.pdf

Suralkar, S.R., A.H.Karode, Priti W. Pawade, 2012. "Texture Image Classification using Support Vector Machine", *International Journal of Computer Technology Application*, Vol 3(1), pp 71-75.

Timo Ahonen, Abdenour Hadid and Matti Pietikainen, 2006. "Face Description with Local Binary Pattern: Application to Face Recognition". Vol. 14, No. 1, December 2016, pp 1-15 IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 28, Issue 12, pp 2037-2041

TSANG TAT MAN project report on: "Image-Based Face Detection System", Oxford Brookes University, pp 1-124. "http://www.peter-lo.com/Teaching/U08096/200912-B1.pdf".

****