International Journal of Image Processing and Vision Science

Volume 1 | Issue 1 Article 7

July 2012

Machine Learning Approach for Smile Detection in Real Time Images

Harsha Yadappanavar
Department of Computer Science and Engineering P.E.S Institute of Technology Bangalore – 560085, Karnataka, INDIA, harshasdm@gmail.com

Shylaja S S
Department of Information Science and Engineering P.E.S Institute of Technology Bangalore- 560085, Karnataka, INDIA, Shylaja.sharath@pes.edu

Machine Learning Approach for Smile Detection in Real Time Images

Harsha Yadappanavar
Department of Computer Science and Engineering
P.E.S Institute of Technology
Bangalore – 560085, Karnataka, INDIA
harshasdm@gmail.com

Shylaja S S
Department of Information Science and Engineering
P.E.S Institute of Technology
Bangalore- 560085, Karnataka, INDIA
Shylaja.sharath@pes.edu

Abstract— Recognizing facial expressions of human beings by a computer is an interesting and challenging problem. A system that performs the operation of face detection and facial feature extraction accurately and in real time would form a big step in achieving a human-like interaction between man and machine. In this paper, we propose a method for detecting Smile in real time Images by machine learning approach. Machine learning method involves training a classifier and using it in real time images to determine smile. Our implemented approach has been tested on several Images from different databases and the achieved results were found to be very satisfactory.

Keywords- Machine Learning, Face Detection, Smile Detection, Haar Classifier, Facial expression analysis

I. INTRODUCTION

The detection of faces and the interpretation of facial expression under varying conditions is an everyday task for humans, which we fulfill without effort. The identity, age, gender as well as the emotional state can be seen from someone's face. The impression we get from a displayed expression will affect our interpretation of the spoken word and even our attitude towards the speaker himself. Humor and sympathy are just two examples for essential information's that are primarily communicated via facial expressions. Hence, they have a high importance for our daily life even though we often are not aware of it.

For computer based systems on the other side, it still is hard to open up this very important channel of communication. The rapidly expanding research in face processing is based on the premise that information about a user's identity, state, and intent can be extracted from images, and that computers can then react accordingly, i.e., by observing a person's facial expression.

Facial expressions are a form of nonverbal communication. They are a primary means of conveying social information among humans. The task of automatic facial expression analysis can be divided into three main steps: face detection, facial feature extraction and classification into expressions.

Face recognition process is influenced by several factors such as shape, reflectance, pose, occlusion and illumination [4]. The face is a highly deformable object, and facial expressions come in a wide variety of possible configurations. Time-varying changes include growth and removal of facial hair, wrinkles and sagging of the skin caused by aging and change in skin color because of exposure to sunlight. Hence, the human face is much more difficult to model and recognize than most industrial parts. This hard challenge is one of the reasons why computer vision research community has been devoted to face recognition for quite some time.

This paper describes the machine learning methodology of detecting one of the facial expression, smile.

The paper is organized as follows. Section 2 describes the alternate methods for smile detection. The methodology used in detecting face region is presented in section 3. Section 4 describes in detail the method for training classifier and method for detecting smile. Finally results obtained are presented in section 5

II. LITERATURE SURVEY

This section deals with the survey of the recent work carried out by researchers in the domain of Facial Expression Analysis. The survey of papers is carried out to know the existing techniques being used for detecting smile in static as well as in real time images.

A. Automatic Analysis of Facial Expressions

Maja Pantic and Leon J.M. Rothkrantz [3] have explored and compared a number of approaches to facial expression detection and classification in static images and image sequences. The investigation compared automatic expression information extraction using facial motion analysis, holistic spatial pattern analysis, and analysis of facial features and their spatial arrangement. Analysis of facial expressions is an intriguing problem which humans solve with quite an apparent ease. They have identified three different but related aspects of the problem: face detection, facial expression information extraction, and facial expression classification. Capability of the human visual system in solving these problems has been discussed. It serves as a reference point for any automatic vision-based system attempting to achieve the same functionality. Among the problems, facial expression classification has been studied most, due to its utility in application domains of human behavior interpretation and HCI. Most of the surveyed systems, however, are based on frontal view images of faces without facial hair and glasses what is unrealistic to expect in these application domains.

B. A survey of Facial Expression analysis and synthesis

Stelios Krinidis, Ioan Buciu and Ioannis Pitas [2] have surveyed the issues associated with facial expressions analysis and synthesis. Analysis of human facial expressions consists of three steps: face detection (tracking), facial feature extraction and facial classification. Automatically classification of human facial expressions is performed according to certain facial action coding schemes, using either spatio-temporal or spatial approaches. An individual specific model is obtained and fitted into the prearranged prototype mesh. Then, the constructed individual facial model is deformed to produce facial expression.

C. Automatic Recognition of smiling and neutral facial expressions

For differentiating smile and neutral facial expression, P.Li, S.L.Phung, A. Bouzerdom and F.H.C Tivive [6] have described architecture for detecting smile from neutral facial expression. The authors have proposed face alignment method to address localization error in existing face detection methods. The smiling and neutral facial expressions are differentiated using a novel neutral architecture. The architecture combines the fixed and adaptive non linear 2D filters. Fixed filter extract primitive features, and adaptive filters are trained to extract complex features. After features are extracted, they are classified by means of linear classifier and classify into smile or neutral.

D. Facial expression recognition

A positive face expression recognition method based on image around the mouth in presented in [7]. Estimation of the direction of face of person is performed first. It has been realized by using particle filter and Fast operator. Next the image around the mouth is detected based on estimated face position. The feature is calculated by Gabor filter and facial expression is recognized by using SVM.

E. Practical Smile Detection

A real time and robust smile detection system is presented in [1] by Jacob Whitehill, Gwen Littlewort, Ian Fasel, Marian Barlett. The paper explores whether current machine learning methods can be used to develop an expression recognition system that operates reliably in more realistic conditions. A new database, GENKI, is presented which contains pictures, photographed by the subjects themselves, from thousands of different people in many different real-world imaging conditions. Results suggest that human-level expression recognition accuracy in real-life illumination conditions is achievable with machine learning technology.

III. PROPOSED METHODOLOGY

A. Face detection

Real time face detection has been performed by using Haar like Feature Classifiers. Haar-like features are rectangular features that can indicate specific characteristics in an image [9]. The idea behind Haar-like features is to recognize objects or features based on the value of simple features, instead of pixel values directly. The Haar-like features have the advantage of very fast computation, because it depends only on the sum of pixels within a rectangle instead of every pixel value. Using an integral image for calculating the sum, one rectangle can be computed with only four references, independent of the size of the feature.

The classification error is represented by:

$$L(y(x), F(x,\theta)) = \begin{cases} 0, & y(x) = sign(w^T x - b) \\ 1, & otherwise \end{cases}$$

Hinge Loss is given by:

$$L(y(x), F(x,\theta)) = max(0, 1 - y(x)F(x,\theta))$$

Total loss shows how good a function (F, θ) is :

$$L(f) = \int_x L(y, F(x))P(x)dx$$

Learning is to find a function to minimize the loss:

$$(F,\theta) = \arg\min_{F,\theta} \int_x L(y, F(x,\theta))P(x)dx$$

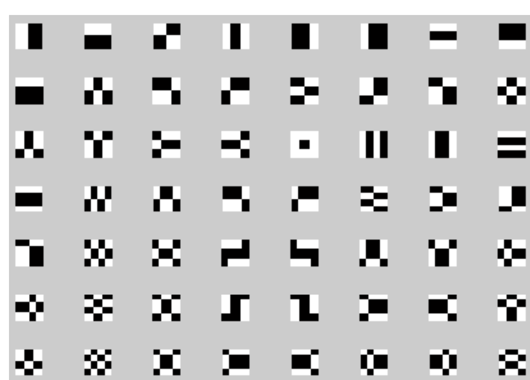

Fig 1: Example of 56 Haar Like Features

The simplest Haar-like feature is a two-rectangle feature. The value of this is calculated as the difference between the sum of the pixels within the two rectangles. This will indicate characteristics like edges or borders between light and dark regions. A three-rectangle feature indicates for instance a dark line or a dark thin area lying between light regions, depending on the size of the middle rectangle. A Four-rectangle feature computes the difference between diagonal pairs of rectangles, and so on. Using the integral image for computing the sum will allow a two-rectangle feature to be calculated with six references to the integral image, a three-rectangle feature with eight references, a four-rectangle feature with nine references.

B. Adaboost Learning Algorithm

The base resolution [4] of the detector is 24x24, which results in a large amount (more features than pixels) of possible features available. A small number of these features can be combined to form an effective classifier. A variant of AdaBoost is used to select this small set of features and is later also used to train the Cascade Classifier.

Fig 2: Features applied to detect face

The first feature selected by AdaBoost seems to focus on the property that the region of the eyes is often darker than the region of the nose and cheeks. The second feature selected relies on the property that the eyes are darker than the bridge of the nose.

Mathematically final classifier is represented by:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

C. Support Vector Machine

The goal of SVM is to separate the data with a hyper plane. The equation of general hyper plane is w'x+b=0, with x being a point (vector), w the weights. The hyper plane should separate the data, so that $w'x_k+b > 0$ for all the x_k of one class, $w'x_j + b < 0$ for all x_j of other class. If the data are in fact separate in this way, there is probably more than one way to do it.

Among the possible hyper planes, SVMs select the one where the distance of the hyper plane from the closest data points (margin) is as large as possible.

Given set of training data

$$\mathcal{D} = \{(\mathbf{x}^1, y^1), \ldots, (\mathbf{x}^l, y^l)\}, x \in \mathbb{R}^n, y \in \{-1, 1\}$$

For $y^i = +1$

$$\mathbf{x}^i \cdot \mathbf{w} + b \geq +1 \quad \text{for } y^i = +1$$

For $y^i = -1$

$$\mathbf{x}^i \cdot \mathbf{w} + b \leq -1 \quad \text{for } y^i = -1$$

Hyper plane is given by

$$y^i(\mathbf{w} \cdot \mathbf{x}^i + b) \geq 1 \qquad \forall i$$

Width of the margin can be classified using

$$M = \frac{(x^+ - x^-) \cdot w}{|w|} = \frac{2}{|w|}$$

Classification function which classifies the data is given by:

$$f(x) = sgn(\mathbf{w}^* \cdot x + b^*)$$

IV. SMILE DETECTION

Before training the classifier, there are two main steps required for object classification: a training set must be constructed for which the true classifications of the objects are known, and a set of object parameters must be chosen that are powerful discriminators for classification. Once a possible classifier has been identified, it is necessary to measure its accuracy.

A. Training Set

Crucial for training the classifier was collection of database of images consisting of smile and non smile images. The images collected had to span a wide range of imaging conditions and also variability in age, gender, ethnicity, facial hair and glasses. Hence meeting all these requirements, we collected around 519 images consisting of 258 smile and 261 non smile faces.

After collection of images, we converted all of them to grayscale and then to standard size of 64 X 64.

B. Feature Selection

Training classifiers is an optimization problem in a many-dimensional space. Increasing the dimensionality of the space by adding more parameters makes the optimization harder (and the difficulty grows exponentially with the number of parameters). There are many facial features which change when a person smiles. In order to optimize the classifier and to lessen the difficulty level, we have chosen only mouth as the potential feature because, mouth plays the major role in identifying whether person is smiling or non-smiling.

Fig 3: Example of face images from database

We segmented the mouth region from all images in database and converted the segmented images to standard size of 16 X 16 making it to 256 feature values or support vectors.

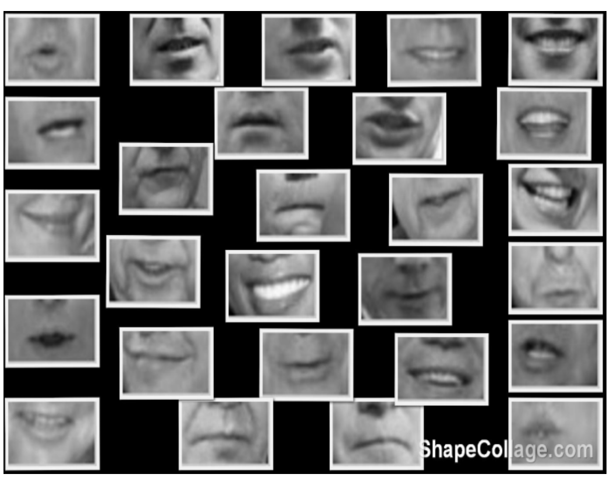

Fig 4: Examples of mouth region from database

C. Design for Detecting Smile

From the database of face images, we extracted the mouth region and then normalized it. These mouth region images are then subjected to training by using Support vector machine. SVM treats the feature values as support vectors and generates a trained classifier model. This trained classifier model can be used to classify new data. In real time, each image frame is considered as a static image. The training has been performed by using 570 face images from the collection of images in database. Following High level diagram describes in brief the methodology used for detecting smile.

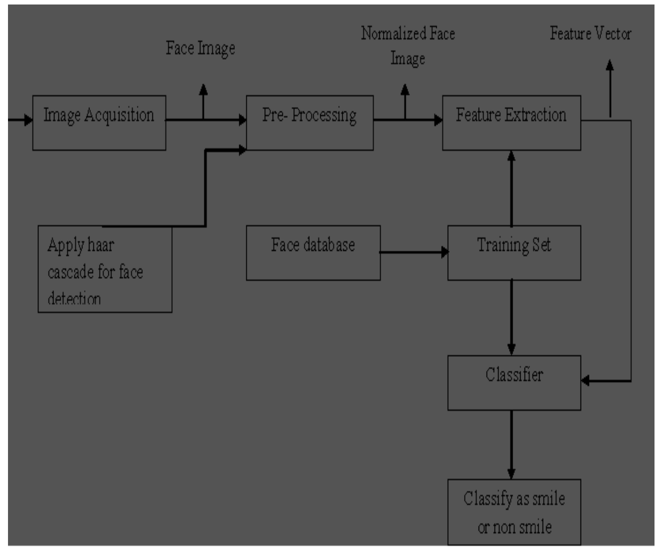

Fig 5 High Level Design of the Proposed Smile Detection system

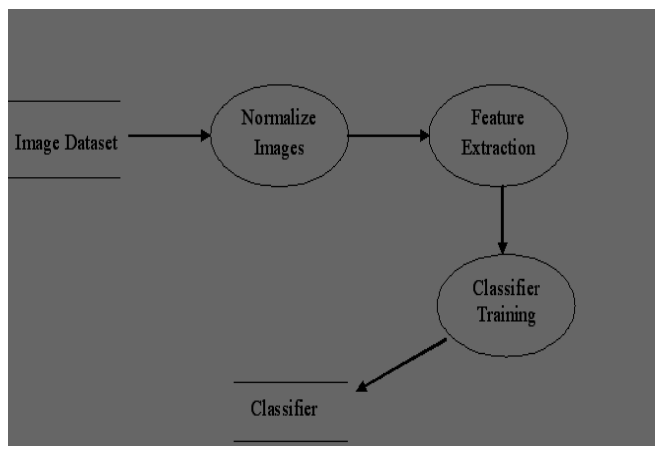

Fig 6 Dataflow diagram for training a classifier

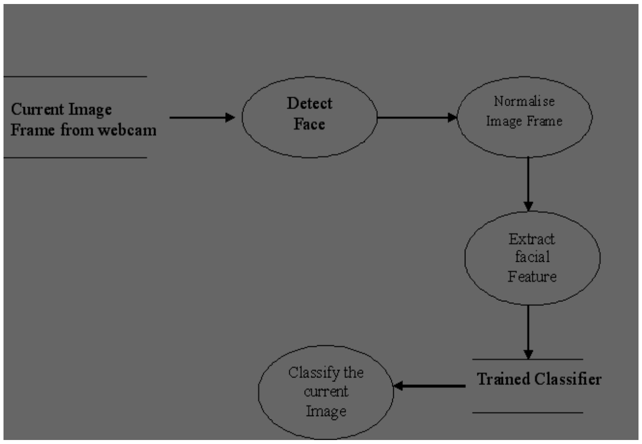

Fig 7: Dataflow diagram for classification of current input image frame

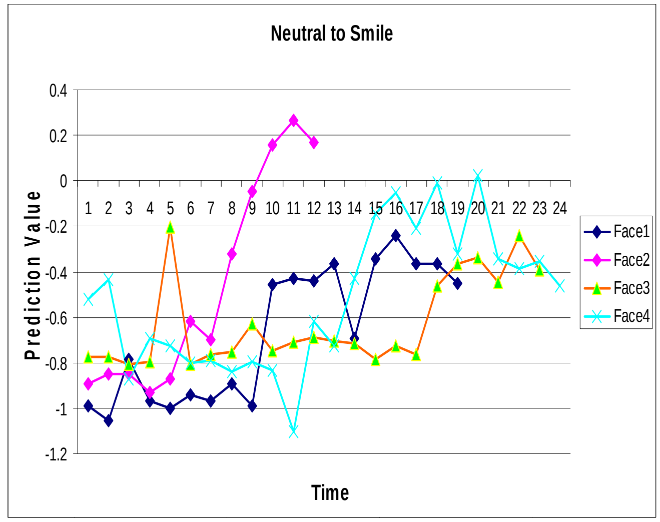

Fig 5: Graph showing values when face changed neutral to smile

V. RESULTS

A. Accuracy of Face Detection

First and foremost, the accuracy of detecting faces in static images was determined. Face detection using Haar-Like Feature classifiers was tested on total of 336 images of different image size. I found out that the overall accuracy almost 99% in face detection with negligible false positives and false negatives. Following table depicts the accuracy.

Table 1: Accuracy of Face Detection approach

Parameters / Total number of Images	Image size	Correct Detections	False Positives	False Negative	Accuracy in %
181	180 X 200	176	2	3	97.2
155	100 X 100	148	5	2	95.4

B. Calculation of threshold

The main task in smile detection approach was to calculate the threshold when the person smile would be detected. Once the classifier is trained, we applied the classifier in real time for around 25 faces and found that the threshold was -0.4 when the person actually smiled. Following graph shows the calculation of threshold of 4 different faces

C. Accuracy of Smile Detection

For calculating the accuracy in smile detection, we applied the methodology for 270 grayscale images of size 64 X 64. We found that out of 270 images, there were 222 correct detections and 36 false positives and 12 false negatives were generated. Thus making it an accuracy of 82.2%. Following table depicts the same.

Table 2: Accuracy of smile detection methodology

Parameters / Total number of Images	Image size	Correct Detections	False Positives	False Negative	Accuracy in %
270	64 X 64	222	36	12	82.2

D. Performance

In measuring the performance, the total number of support vectors used in training and in real time is selected. I noticed that as the feature size decreases, the time taken to detect smile also decreases. Following chart shows the time taken to detect smile when the corresponding support vector is used.

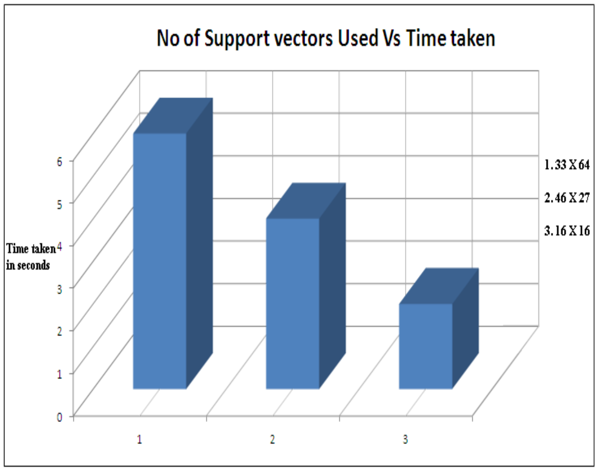

Fig 6: Chart showing Time taken to detect smile when the corresponding support vectors are used

E. Comparison with existing smile detection methods

Finally, I compared the time taken by our implemented smile detection approach with Sony camera's smile shutter function and Samsung Galaxy mini S5570. Both the Sony camera's smile shutter function and Samsung Galaxy's smile shutter are only capable of detecting larger smile, i.e. when the person largely smiles, and are not capable of detecting slight smile. The comparison was performed for 10 different faces.

Table 3: Comparison of our algorithm with other smile detection methods

Camera / Parameters	Sony Smile Shutter	Samsung Galaxy mini S5570	Our Implemented Smile Detection algorithm
Time Taken(in secs)	2.5	2	2
Slight Smile	No	No	YES

VI. SUMMARY AND CONCLUSIONS

Machine learning approach produces the best performance for facial expression analysis. Hence it is possible to produce accurate facial expression analysis by machine learning approaches.

The current dataset used in the expression recognition are too small and lack variability in imaging conditions. Current machine learning methods may require on the order of 1000 to 10,000 images per target facial expression. So, more the number of training samples more will be the accuracy.

The above method has been implemented in OpenCV platform and takes approximately 2 seconds to detect smile and achieves around 90% accuracy.

ACKNOWLEDGMENT

The authors would like to extend their sincere gratitude to PES Management and all faculty members and friends in all support provided for carrying out this research work.

REFERENCES

[1] Jacob Whitehill, Gwen Littlewort, Ian Fasel,Marian Bartlett "Toward Practical smile Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2106 – 2111, Vol 31, No 11, 2009

[2] Stelios Krinidis, Ioan Buciu and Ioannis Pitas, "Facial expression analysis and synthesis: A survey", *A white paper by Department of Informatics*, Aristotle University of Thessaloniki

[3] Maja Pantic and and Leon J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 22, no 12, pp 1424 - 1445 2000

[4] Paul Viola and Michael Jones "Robust Real Time Object Detection", *Second International Workshop On Statistical and Computational Theories Of Vision Modeling, Learning, Computing, and Sampling*, 2001

[5] Wataru Hirata, Joo Kooi Tan, Hyongseop Kim, Seiji Ishikawa, "Recognizing Facial Expression for Man-machine Interaction", *ICROS-SICE International Joint Conference 2009* August 18-21, 2009, Fukuoka International Congress Center, Japan

[6] P Li, S.L.Phung, A. Bouzerdom and F.H.C Tivive "Automatic Recognition of Smiling and Neutral Facial Expressions", *Digital Image Computing: Techniques and Applications*, pp. 581 – 586, 2010

[7] Yu-Hao Huang and Chiou-Shann Fuh "Face Detection and Smile Detection", *White Paper*, 2009

[8] I M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, 2006

[9] N Phillip Ian Wilson and John Fernandez "Facial feature detection using Haar classifiers", *Journal of Computing Sciences*, Volume 21 Issue 4, 2006

[10] http://www.cs.ucla.edu/~falaki/pub/adaboost.pdf

[11] http://opencv.willowgarage.com/wiki/