Multiview Smile Detection by Gabor Wavelet Decision Tree Classifier

J Shekaina
Head of Department
Dept of Electronics & Communication Engineering
C.S.I Institute of Technology, Tamilnadu, India

Priya J L
M.E Scholar
Dept of Electronics & Communication Engineering
C.S.I Institute of Technology, Tamilnadu, India

ABSTRACT

This paper proposes a novel and efficient methodology for the detection of smile in face images captured in real-world scenarios. It is a challenging research area and possesses so many potential applications in new interface for entertainment systems, expression recognition technologies, social robots etc. This project proposes an efficient approach to multiview smile detection, in which the intensity differences between pixels in the grey scale face images are used as simple features. Gabor wavelet features in which facial expressions are analysed using a multi-orientation set of Gabor wavelets which are in alignment with the face is adopted in this work. Decision tree Classifier is chosen for classification of smile and non-smile faces. The experimental results illustrates that the algorithm achieves very good performance on the database, exceeding or competing with the state-of-art methods. The proposed method is compared with the existing methods and this method is efficient and has good accuracy.

General Terms

Smile Detection, Feature Extraction

Keywords

Gabor wavelet Feature Extraction, Decision tree classifier.

1. INTRODUCTION

Smiling is a very common thing in people's daily life, but almost never considers the power of it. It is the most easily recognizable facial expression that occurs among human beings. Smile is understood by everyone-despite religion, culture, race or nation and often indicates happiness, appreciation, pleasure or satisfaction. Smile can be spontaneous or artificial. According to cross cultural studies, smile is a means of communication throughout the world. During smiling 5 to 53 facial muscles are involving. Developmentally, smiles are one of the first emotion expressions to appear, they occur with relatively high frequency throughout the lifespan, and they can express a variety of meanings, including joy, appeasement, and greetings, and they often serve to mask anger, disgust and other negative emotions. Facial expression is one of the most powerful and immediate means for humans to communicate their emotions, cognitive states, intentions, and opinions to each other. Facial expressions play a critical role in the evolution of societies, helping coordinate social interaction, promoting group cohesion, and maintaining social affiliations. Significant applications of expression recognition technology include diagnosis and monitoring of clinical disorders,

Fig. 1. Examples of real life smile faces (top two rows and non smile faces (bottom two rows), from the GENKI4K database

tutoring systems that are sensitive to the expression of their students, evaluation of behavioral and pharmacological treatments, new interfaces for entertainment systems, smart digital cameras, and social robots.

In this paper, smile detection in facial images captured from real world scenarios is focused. Fig.1 shows some examples collected from real world situations. An efficient and comprehensive approach to smile detection, using the intensity differences between pixels and Gabor wavelet features in the greyscale face images as simple features is presented. Decision tree classifier is employed for classification. Experimental results depicts that this approach achieves higher accuracy than the state-of-the-art method and is significantly faster.

In the last two decades the machine analysis of facial expressions has been an active research topic. Most of the current works have been focused on analyzing a set of prototypic emotional facial expressions. These works utilized the data, collected by asking subjects to pose deliberately these expressions. But in real life situations the exaggerated

facial expressions occur rarely. Spontaneous facial expressions induced in natural environments differ from posed expressions, i.e., both in terms of which facial muscles move and how they move dynamically. Cohn and Schmidt [1] observed that spontaneous smiles are of lesser amplitude and have a long consistent relationship between amplitude and duration than posed smiles. Recently, realistic problem of analyzing spontaneous facial expressions grabbed much research attention. To capture the complex decision boundary among spontaneous expressions is indeed difficult.

The rest of this paper is organized as follows. First the experiments, surveys and other related works are reviewed briefly in Section 2. The experimental procedures and database used and performance evaluations in the proposed method is presented in detail in Section 3. Finally Section 4 gives the conclusions of this work. The future works in this area are given in section 5.

2. LITERATURE REVIEW

Although there is a vast amount of literature and works on facial expression recognition, few papers have focused exclusively on smile detection. Caifeng Shan [2] proposed a smile detection system in which pixel intensity differences are used as features. AdaBoost was adopted to choose and combine weak classifiers based on intensity differences to form a strong classifier in this system. This approach attains 85% accuracy by examining 20 pairs of pixels and 88% accuracy with 100 pairs of pixels. The two major approaches on face image analysis are local feature-based and image-vector-based. Y. Shinohara et.al [3] proposes a hybrid of these two approaches. This method used higher order Local Auto-Correlation (HLAC) features and Fisher weight maps. Experimental results summarized that the recognition rate of this method using the Fisher weight map (FWM) and HLAC features was 97.9%, while Fisher faces method was 93.8% and HLAC without a weight map was 72.9%. The hybrid method outperformed the Fisher faces method and the HLAC-features-based method. This method achieved high accuracy, but only limited data were used in this study. A. Ito et.al [4] describes a method to detect smiles and laughter sounds from the video of natural dialogue.In this paper, a six-dimension feature vector is extracted to describe the lip and cheeks, which incorporates a perceptron classifier for smile detection. Testing done on three video sequences, this method achieves the accuracy of 60%−85%. U. Kowalik et.al [5] explains BROAFERENCE, a test bud for studying future oriented multimedia services and applications in distributed environments. In feature extraction, a 16-dimension feature vector derived by tracking eight mouth points is used with a neural network classifier to detect smiles. However, performance evaluation was absent. O. Deniz et.al [6] described both face detector and smile detector for PUIs. A smile detector based on the Viola–Jones cascade classifier is described. Training done on 5812 images (i.e., 2436 positive and 3376 negative images), the detector achieves the accuracy of 96.1% on 4928 testing images. Even then, the used face images are mainly frontal, with limited imaging conditions. Whitehill et al. [7] presented a comprehensive study on practical smile detection. They collected the GENKI database comprising of 63,000 real-life face images from the Web. They investigated different parameters, including size and type of data sets, image registration, facial representation, and machine learning algorithms. Their study suggested that high detection accuracy is achievable in real-life situations, with the argument that the order of 1000–10,000 images that have a wide range of imaging conditions and personal variables is
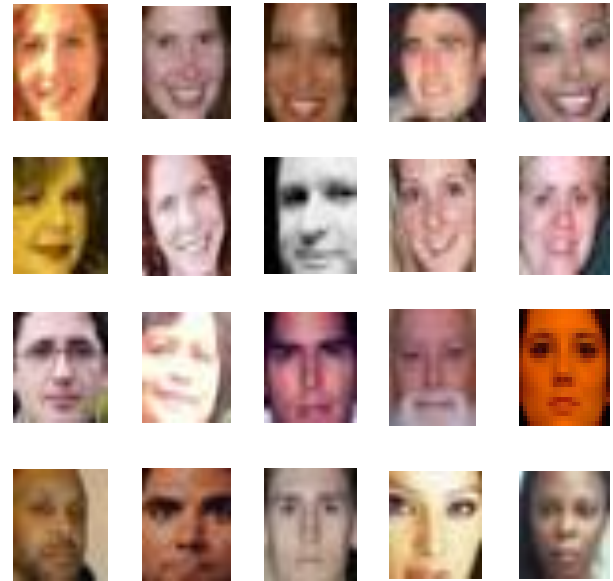

Fig. 2. Normalized face images: smile (top two rows) and no smile (bottom two rows)

required for training. C. Shan et.al [8] empirically evaluated facial representation based on statistical local features, Local Binary Patterns, for person-independent facial expression recognition. Different machine learning methods, including template matching, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and the linear programming technique, to perform facial expression recognition using LBP features are examined. LBP features perform stably and robustly over a range of low resolution face images. Best recognition performance is obtained by using SVM with Boosted LBP features. N. Sebe et.al [9] presented an authentic facial expression database. This paper integrated the classifiers and a face tracking system to build a real time facial expression recognition system.

3. PROPOSED SYSTEM

3.1 Database

The experiments are conducted on the publicly available GENKI4K database from the Web. GENKI4K is a database of images released by Machine Perception Laboratory (MP Lab). The database consists of 4000 images (2162 "smile" and 1828 "no smile"); some samples are shown in Fig. 1. The images in the database span a wide range of subjects, illumination, facial appearance, geographical locations and camera models. The imaging conditions include both outdoors and indoors, as well as variability in age, gender, imaging situations, ethnicity, facial hair, and glasses [7]. All images are labelled for smile content (1=smile, 0=no smile). The ground truth on "smile" and "no smile" is provided with the database.

In this work, the images are converted to greyscale. The face images are normalized to reach a canonical face of 24 x 24 pixels. Fig.2 illustrates the normalized faces. Among the images 340 images are selected as training set with similar number of "smile" and "no smile" samples. Then a cross-validation procedure is adopted using the testing set.

3.2 Boosting Pixel Differences

A vital step in facial expression analysis is extracting significant and effective features from original face images. Geometric features are exploited based on the feature points

(e.g.eye points, mouth corners). However reliable and accurate detection and tracking of feature points are difficult to capture in real-world unconstrained scenarios. Another category of features called appearance features take into consideration the appearance (skin texture) of the face. They are less sensitive to errors in feature point detection and these features encode changes in skin texture that are important for facial expression modeling. Hence appearance features seems more promising for unconstrained facial expression analysis. As an example, the responses of Gabor filters at multiple spatial scales, orientations, and locations have been used traditionally [10] and have been proven successful for smile detection [7]. Other features include the Haar wavelet (or boxlet) features [12], HLAC features [3], Local Binary Pattern [8] etc.

A key concern in practical applications is speed or computational efficiency. Since the computational resources are limited, it is highly desired that the features used can be computed easily and efficiently. Baluja *et al.* [13], [14] described to use the relationship between two pixels' intensities as features. The work obtained high accuracy on face orientation discrimination and gender classification by comparing the intensities of a few pixels.

In this paper, the relationship between two pixels' intensities as image features is examined. It is proposed to use the intensity difference between two pixels as a simple feature. In intensity difference procedure, for each pixel of input image the intensity difference between the pixel under consideration and the remaining pixels is evaluated. It is observed that the intensity difference covers vast information about the facial appearance during the time of a facial expression. To visualize the features that have been extracted, the weight for each pixel is accumulated, and the greyscale intensity is proportional to the times of that pixel being used. It is clear that the involved pixels are distributed mainly in the regions around mouth, with a few from the eye areas. The major difference between smile and no smile faces surrounds in the mouth region and the eyes, where smile faces have their mouth open, whereas no smile faces have mouth closed.

3.3 Gabor wavelet Features

In [7], Gabor features achieve superior performance on smile detection. Therefore, Gabor features are considered in this paper. To achieve more accuracy Gabor wavelet features are used along with pixel intensity features. Gabor functions provide the optimal resolution in both the time (spatial) and pixel domains among various wavelet bases. They have been successfully applied for smile detection due to the robustness of Gabor features against local distortions caused by variance of illumination, expression and pose. The simple cells of the visual cortex of mammalian brains are best modeled as a family of self-similar 2D Gabor wavelets. These features consist of a dictionary of frequency scaled and rotated Gabor functions. They extract directional features from images. Gabor wavelet transform has both the multi-resolution and multi- orientation properties and are optimal for measuring local spatial frequencies. The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor wavelets.

3.4 Illumination Normalization

Researchers have tried various techniques to deal with illumination variation [11]. As observed in Figs. 1 and 2,
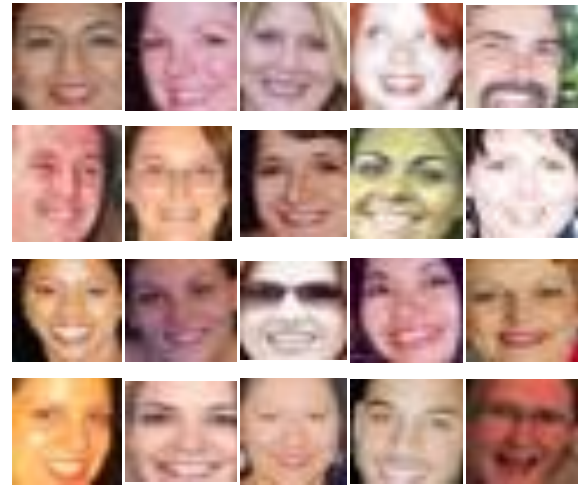

Fig. 3. Examples of correct detection

illumination variation is one of the difficulties for smile detection in real-life faces. In this work, for handling illumination variation Histogram Equalization (HE) is employed. HE is a simple and widely used technique for normalizing illumination effects by the technique of contrast enhancement. Normally obtained photographs with under-exposure or over-exposure due to uncontrolled environment lightness, makes it difficult to recognize the details of images. This method increases the global contrast of many images, especially when the important data of the image is represented by close contrast values. It allows for areas of lower local contrast to gain a higher contrast. Histogram equalization achieves this, by effectively spreading out the most frequent intensity values. However, if the image is first converted to another color space, Lab color space, then the algorithm can be applied to the luminance or value channel without resulting in changes to the hue and saturation of the image.

3.5 Decision Tree Classifier

Classification, which is the task of assigning features to one of several predefined categories, is a major problem that encompasses many diverse applications. In the conventional methods Adaboost classifier [2] that combine weak classifiers based on pixel differences to form a strong classifier for smile detection, Neural network classifiers [4] that makes prediction based on a linear predictor function, Support Vector machine classifier [7] were employed.

In this work Decision Tree Classifier is chosen to classify the features and frame a model that lead to the conclusion of "smile''and "no smile". The goal of decision tree is to create a the value of a target variable based on several input variables. Decision Tree classifier is chosen since it is simple to understand and interpret, requires little data preparation and it performs well with large data in a short time. This tree has three types of nodes: root node, internal node and leaf node. Root node has no incoming edges and zero or more outgoing edges. Internal node has one incoming edge and two or more outgoing edge. Leaf node has one incoming edge and no outgoing edge. Each leaf node is assigned a class label. For classification, first a training set comprising of features whose class labels are known is provided. The training set is utilized
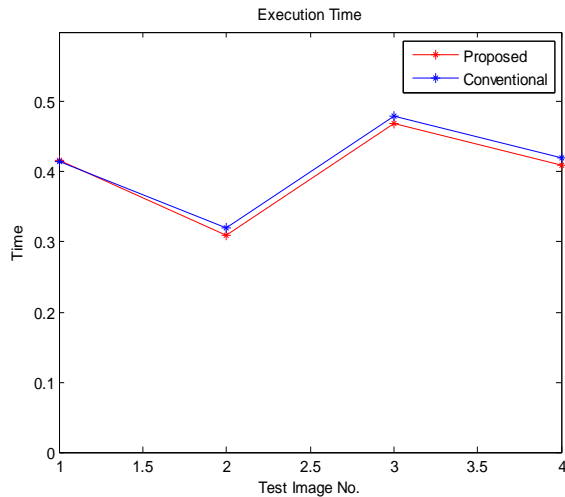
Fig. 4. Execution time as a function of database image sets

to build a classification model, which is further applied to the test set, consists of features with unknown class labels. The classification threshold for each feature in an image is estimated during the training phase. A series of conditions and there results are organized in the form of a decision tree. This decision tree helps to visualize and interpret the results.

3.6 Performance Evaluation

Fig.3 shows some examples of correct smile detection. The Execution time and detection accuracy employing both Gabor features and pixel intensity difference features and pixel intensity features alone are compared. Given a new test image, the first method extracts Gabor features and then performs classification with the Decision Tree classifier. Smile detection function often requires extensive training and performance is dependent on training data. This experiment, is trained on 340 face images. For the second method, intensity differences alone of a limited number of pixel pairs from the given image are computed. Fig.4 depicts a comparison of execution time between proposed method and the conventional one. Compared to the feature extraction by pixel intensity features alone, it can be observed that the proposed detector is significantly faster. Even with 500 features, the experiment demonstrated that the proposed method is indeed rapider. Fig.5 illustrates a comparative study about the detection accuracy of the two methods of smile detection. The detection accuracy achieved by proposed method is higher than 90% compared to the conventional one. Thus the comprehensive approach of pixel intensity difference features and Gabor wavelet features possess a robust and efficient methodology in smile detection compared to the state-of-art methods.

4. CONCLUSIONS

This paper first reviewed the existing techniques in smile detection providing a comprehensive survey, critical challenges and possible future research directions towards practical smile detection. Many insights are derived from the above discussions. Our method was designed to work well with challenging data and achieved a significantly high detection rate. In this paper, the integration of the intensity differences between pixels and Gabor wavelet features in the greyscale face images, utilized for feature extraction are investigated. This work adopts Decision tree classifier for
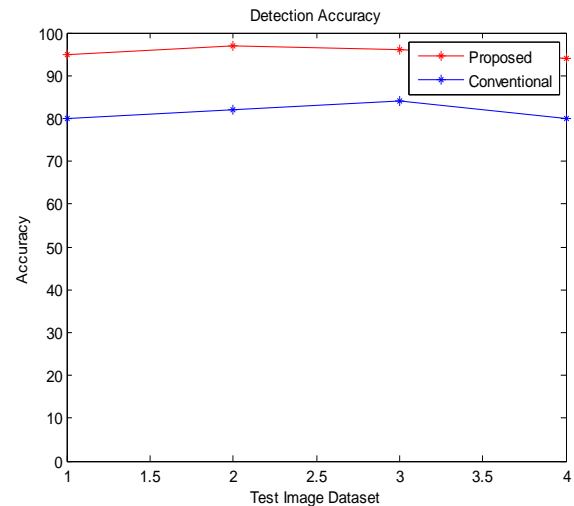

Fig. 5. Detection accuracy as a function of test image data set

classification of smile and no smile images. The proposed system performs robust recognition of smiles from images taken both indoors and outdoors, using a gallery of frontal views of the subject's faces. Computational efficiency and the capability to deal with pose variation to some extend without any explicit technique are the advantages of this system. Experiments illustrate that this approach has an accuracy that exceeds 90% compared to the state-of-the-art methods.

5. FUTURE WORKS

A largely unexplored area of research is that of context dependency. The interpretation of human smiles is often context dependent. For example, a smile can be a display of humbleness, irony, greeting, joy or masking anger. To interpret a smile it is important to have a clear idea of context in which it is displayed, i.e., who the expresser is, where the expresser is, who the receiver is etc.

Various applications using smile detection can be envisaged in the near future, boosting further interest in doing research in the fields of smile recognition and interpretation. Non-verbal information transmitted by facial expressions is of great importance in different areas, including image interpretation, psychological studies, more engaging human-machine interfaces, robotics and virtual reality.

6. REFERENCES

[1] J. F. Cohn and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles", *Int. J.Wavelets, Multiresolution Inf. Process.*, vol. 2, pp. 1–12, 2004

[2] Caifeng Shan,"Smile Detection by Boosting Pixel Differences"*, IEEE transactions on image processing,* vol. 21, no. 1, pp 431-436, 2012

[3] Y. Shinohara and N. Otsu, "Facial expression recognition using fisher weight maps," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog*, 2004, pp. 499–504.

[4] A. Ito, X.Wang, M. Suzuki, and S. Makino (2005), "Smile and laughter recognition using speech processing and face recognition from conversation video," in *Proc. Int. Conf. Cyberworlds*, 2005, pp. 437–444.

[5] U. Kowalik, T. Aoki, and H. Yasuda, "Broaference—a next generation multimedia terminal providing direct feedback on audience's satisfaction level," in *Proc. IFIP*

TC13 Int. Conf. Human-Comput. Interact. (INTERACT), 2005, pp. 974–977.

[6] O. Deniz, M. Castrillon, J. Lorenzo, L. Anton, and G. Bueno*,* "Smile detection for user interfaces," in *Proc. Int. Symp. Adv. Vis. Comput.*, 2008, pp. 602–611

[7] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Towards practical smile detection," *IEEE Trans. Pattern Anal. Mach. Intell.,*vol. 31, no. 11, pp. 2106–2111, Nov. 2009

[8] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis.Comput.,* vol. 27, no. 6, pp. 803–816, May 2009

[9] N. Sebe, M. S. Lew, I. Cohen, Y. Sun, T. Gevers, and T. S. Huang*,* "Authentic facial expression analysis," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog,* 2004, pp. 517–522

[10] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 974–989, Oct. 1999

[11] X. Tan and B. Triggs*,* "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures,* 2007, pp. 168–182.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 511–518.

[13] S. Baluja, M. Sahami, and H. Rowley, "Efficient face orientation discrimination," in *Proc. IEEE Int. Conf. Image Process.*, 2004, pp.589–592

[14] S. Baluja and H. A. Rowley, "Boosting set identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, 2007