ANALYSIS OF FACE RECOGNITION SYSTEM WITH FACIAL EXPRESSION USING CONVOLUTIONAL NEURAL NETWORK AND EXTRACTED GEOMETRIC FEATURES

C.R VIMALCHAND

ABSTRACT

Faces contain a great deal of valuable information in automatic image processing to extract semantic content for effective analysis and pattern recognition. It requires efficient algorithms that are capable to extract relevant semantic information. Recognizing a person on a photograph tells a lot about the overall content of the picture. The principal aim of facial analysis is to extract valuable information from face images such as position in the image, facial characteristics, facial expressions, the person's gender or identity. Now we consider the most important approaches to facial image analysis and present novel methods based on convolutional neural networks to detect, normalize, and recognize faces and facial features. In this paper the problems of automatic appearance-based facial analysis with machine learning techniques are described and common specific sub-problem like facial feature detection, analyzing geometric features and face recognition which are crucial parts of many applications in the context of indexation, surveillance, access-control or human machine-interaction are handled. To tackle these problems a new technique called *Convolutional Neural Network* (CNN) is used which is inspired by biological evidence found in the visual cortex of mammalian brains and can be applied to many different classification problems. The proposed system is a CNN- based method for *automatic facial feature detection and extraction* that employs a hierarchical procedure which first detects the eyes, the nose and the mouth and then refines the result by identifying 10 different facial feature points. The detection rate is 87% to 96% for different databases tolerating an error of 10% of the inter-ocular distance. Finally a novel face recognition system based on CNN architecture learning a non-linear mapping of the image space into a lower dimensional sub-space where the different classes are more easily separable. This approach produces better result for different facial databases in detection and recognition with classical face recognition approaches using PCA or LDA.Back propagation algorithm is used as learning algorithm for CNN which is a Multi-Layer Perceptrons. A *Siamese CNN* is used for face verification by receiving two face images and comparing to decide if they are same. .

Index Terms

Human-machine interaction, Convolutional neural network, automatic facial feature detection, Automatic facial feature extraction, inter-ocular distance, Siamese CNN etc.

1. Introduction

Machine recognition of faces is becoming important due to its wide range of commercial and law enforcement applications, which include forensic identification, access control, border surveillance and human computer interactions. Many techniques are available which apply face recognition such as stastical projection method, Hidden Markov Models, Support Vector Machines and Neural Networks. The most common and effective approach for face recognition is based on artificial neural networks. Here one particular approach based on convolutional Neural Network is taken for *face alignment* and *facial feature detection* which shows to be crucial for many facial analysis applications. A Multi-layered feed forward NN is a powerful machine learning technique as they can be trained to approximate complex non-linear functions from high dimensional input examples. Standard Multi-layer Perceptrons (MLP) are used in pattern recognition systems to classify signatures from a separate feature extraction algorithm operating on the input data. The problem with this approach is that when the input dimension is high, as in images, the number of connections, thus the number of free parameters also high. So each of the hidden nodes are fully connected to the input layer and another disadvantage of this type of MLP comes that its input layer has fixed size and the input patterns have to be presented well aligned and/or normalized to this input window which makes it a complex task. The Convolutional Neural Networks are an approach that tries to handle these problems by automatically learn local feature extractors; they are invariant to small translations and distortions in the input pattern. They implement the principle of weight sharing which drastically reduces the number of free parameters and thus increase their generalization capacity compared to other NN.

1.1 Convolutional Neural Network model and Neocognitron

The first implementation of a CNN was the so-called *Neocognitron* proposed by *Fukushima* which has been originally applied for the problem of handwritten digit recognition. The neocognitron makes use of perceptive fields, i.e. each neuron is only connected to a sub-region corresponding to a certain number of neighboring neurons, in the preceding layer. Local connections have been used many times with NNs and they can be used to detect elementary visual features in images such as oriented edges, end points or corners. Lecun's convolutional neural network was trained by Back propagation algorithm and applied to the problem of handwritten digit recognition.

In the world of computers and recognition it is not develop an automatic system to recognize the identification of faces and faces expression. In natural interpretation, the humans can easily detect the expressions the indication of faces and facial expressions. There are many challenges in face expression recognition such as (i) face detection and segmentation from a captured image (ii) extracting the facial expression information ans (iii) the process of classification of the face expression in emotional state. Generally there are 3 major steps in facial expression recognition. The first step is to detect the face. The second step is to extract the facial expression information (facial features) that convening the facial expression the third step is to classify the facial display conveyed by the face. The features are given as input to CNN for classification. The usually extracted facial features are geometric features such as the face of facial components (eyes, mouth etc.) and the locations of facial characteristic points (corners of the eyes, mouth etc.) or appearance features representing the texture of the facial skin in specific facial areas including wrinkles, bulges and furrows. The perceptrons of *Siamese CNN* are trained to recognize the facial expression of the same person.

2. Related work

The pain and happy expressions are important facial expressions. Md.Monwar et al use location and shape features to represent the pain information. These features are used as the inputs for standard back propagation algorithm in the form of three-layer neural network, one hidden layer for classification of painful and painless faces. This approach can be adopted for happy expressions. The Facial Action Coding System (FACS) (Ryan, Cohn, Lucey, Saragih, Lucey, Torre, & Rossi, 2009) and (Lucey, Cohn, Lucey, Matthews, Sridharan & Prkachin, 2009) is currently the most widely used method in recognizing facial expressions. FACS encodes the contraction of each facial muscle (stand alone as well as in combination with other muscles) changes the appearance of face. It has been used widely for the measurement of shown emotions.

Gizatdinova and Surakka used feature-based method for detecting landmarks from facial images (Gizatdinova, & Surakka, 2006). The method was based on extracting oriented edges and constructing edge maps at two resolution levels. Edge regions with characteristicedge pattern formed landmark candidates. Cootes et. al and Ratliff et. al developed the Active Appearance Model AAM that shown strong potential in a variety of facial recognition technologies (Ratliff, Patterson, 2008) and (Ratliff, & Patterson, 2008).

The first implementation of a CNN was the so-called Neocognitron proposed by Fukushima which has been originally applied to the problem of handwritten digit recognition. The neocognitron makes use of receptive fields i.e each neuron is only connected to a sub-region corresponding to a certain number of neighboring neurons, in the preceding layer. This idea has been inspired by the discovery of locally-sensitive, orientation-selective neurons in the cat's visual system by Hubel and Wiesel. Local connections have been used many times with NNs. They can be used to detect elementary visual features in images, such as oriented edges, end points or corners. As one feature extractor can be useful in several parts of the input image, the weights are forced to be identical for all possible locations of the receptive field, a principal called weight sharing.

An important breakthrough of CNNs came with the widespread use of the Back-propagation learning algorithm for multi-layer feed-forward NNs. LeCun et al. presented the first CNN that was trained by Back propagation and applied it to the problem of handwritten digit recognition.

Siamese Neural Networks have first been presented by Bromley et al. using Time Delay Neural Networks (TDNN) and applying them to the problem of signature verification, i.e. to verify the authenticity of signatures. This idea was then adopted by Chopra et al. who used Siamese CNNs and employed them in the context of face verification. More precisely, the system receives two face images and has to decide if they belong to the same person or not.

3. Proposed System

This paper propose a system for classifying the six basic emotions (anger, disgust, fear, happy, sad, surprise) in addition to the neutral one using two types of features. It generally enters the image that contains the face under check into face detection process to segment the face image. Then the feature extraction process will applied on the face image to produce a feature vector that consists of two types of features: geometric features and appearance features which represents a pattern for facial expression classes. Finally this feature vector used as an input into the radial basis function for CNN to recognize the facial expression. The block diagram of the proposed system is shown in Figure1.

3.1 Facial Features Extraction

Generally the most important step in the field of facial expression recognition is the facial feature extraction which based on finding a set of features that conveying the facial expression information. This problem can be viewed as a dimensionality reduction problem (transforming the input data into a reduced representation set of features which encode the relevant information from the input data). This paper use two methods to extract the facial features: geometric features and appearance features.

3.1.1 Geometric Features Extraction

In this step 19 features are extracted from the face image. First; the segmentation process is performed to divide the face image into three regions of interest: mouth, nose and two eyes and two eyebrows. This is done by taking into consideration that the detected face is frontalor near frontal and assuming certain geometric constraints such as: position inside the face, size and symmetry to the facial symmetry axis. Second; the facial characteristic points (FCPs) are located in each face component using mouth, nose, eyes and eyebrows FCPs extraction techniques. Finally; certain lengths between FCPs are calculated using Euclidean distance D:

$$D= \sqrt{(x1-x2)^2+(y1-y2)^2} \qquad (1)$$

Where (x1, y1) and (x2, y2) are the coordinates of any two FCPs P1(x1, y1) and P2(x2, y2) respectively. Also two angles in the mouth area are computed to represent with geometric lengths the geometric features.

3.1.2 Mouth FCPs Extraction:

The first process after face detection is the segmentation process. The mouth area chosen by taking the portion of the face image from 25% of the face image width after the left boarder to 25% of the face image width before the right edge and from 66.67% of the face image height after the top boarder to 5% of the face image height before the bottom boarder. The next step is detecting the FCPs inside the mouth region. This is done by using two transformations of the mouth image after applying "Spatial sharpening filter "to sharpen the mouth image for more clarification of mouth details. The first one is transform the mouth image from gray scale into binary image using certain threshold τ which chosen as a function of the entire mouth image threshold **T** and factor † as the following equation:

$$\boldsymbol{\tau = T / \dagger} \qquad (2)$$

Where the value of † determined by manual iterations (empirically). Due to the large diversity in mouth shape, It is recoded that best results of **†=3**. Binary dilate morphological operation with linear $1 X 5$ structural element and erode morphological operation with linear $5 X 1$ structural element are applied after the first transformation.

The second transformation is converting the mouth image into binary using the canny edge detection with also static threshold value which after manual iterations chosen to be 0.5. Also close morphological operation with structural element $5 X 1$ is applied after the second transformation. The two results are added together to get the final binary image of the mouth.

The previous transformations have overcome the problem of mouth shape diversity. In case of angry, disgust, normal, sad and surprise classes the first transformation can get the mouth region in good way because of the mouth often represented by the two lips only unlike in the fear and happy classes where the teeth often appears and results to generate many objects after the transformations. So using canny edge detection is more accurate in the last category of classes. Since the first transformation is robust for the first category of classes and the second for the second; the proposed approach here is using the both to get an object which represents the mouth region in the mouth image. After that the labeling operation is performing then using the blobs analysis to determine the area and "Bounding Box" properties of each object in the image. The object with maximum area selected to be the mouth rather than other candidate objects. Using the bounding box property of the mouth object will be easy to determine each of the four mouth FCPs as a mean point of none zero values of the corresponding boarder (left, right, top and bottom).

3.1.3 Nose FCPs Extraction:

The nose FCPs are mainly used to measure the distance from them to eyes centers and mouth FCPs. Also the first step is the segmentation process with the values 40%, 40%, 45% and 30% for the left, right, top and bottom boarders respectively to get the nose image. Unlike the mouth in varying their shapes with respect to facial expression classes; the nose hole shape often have the same shape as circle (hole). So an iteration process is performed to get at least two holes from the binary image and the maximum two object are selected after labeling and blob analysis operations. After that the "Centroid" property of the two holes are determined to locate the corresponding two noses FCPs.

3.2.1.3 Eyes and Eyebrows FCPs Extraction:

In eyes and eyebrows the proposed approach is to deal with each eye-eyebrow separately. Also starting with the segmentation process with the values 20%, 20%, 20% and 50% for the left, right, top and bottom boarders respectively is the first step to obtain eyes and eyebrows area. Taking into account the symmetry of the frontal face image the left eye-eyebrow pair is separated from the right one by taking the left half of the segmented area for it and the right one for the other. The next step is to separate each eye-eyebrow to deal with it singly. This is done by finding a separated border between them. The integral projection method is used to determine the horizontal borders of the line between the lips. The boarder finding algorithm consists of five steps:

1. Apply "Prewitt" filter that emphasizes horizontal edges by approximating a vertical gradient.

2. Compute the vertical integral projection $s(x)$ for $n\ X\ m$ eye-eyebrow image by using equation (3):

$$s(x) = \sum_{x=1}^{m} \boldsymbol{l(x,y)}$$

(3)

Where $x \in \{1, 2, \ldots.. n\}$ is the row number.

3. Smooth the results using a moving average filter to minimize the number of peaks around global maxima.

4. Find global maxima

5. If the number of maxima is greater than two peaks; the border is computed as the average position of the maximum two peaks position. Else it is one peak and then the border will be the position of this peak.

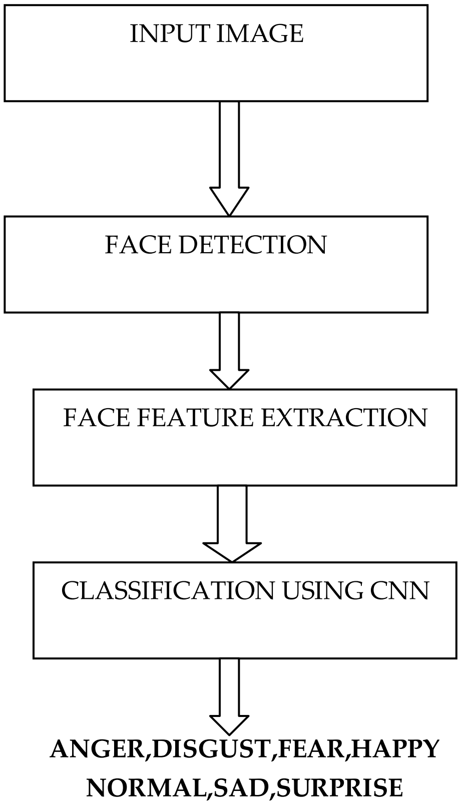
The next step after determining the eye-eyebrow border is to separate the eye and eyebrow using this border. Then the FCPs of the eye and eyebrow can be determined using the transformation to binary with threshold technique. Also some image enhancements are applied using Gaussian filter and image histogram equalization before getting the binary image. And also "Close" morphological operation (with "Disk" structural element in eye image and "rectangle" structural element in brow image) and labeling operation are applied in eye image. Finally, by using blobs analysis the area property is calculated to choose the object with the maximum area that corresponds to the eye. Then locate the eye center from the "Centroid" property in eye image. After that the "Bounding Box" property is used to determine the object with maximum area that corresponding to the eyebrow. Then left, top and right FCPs in eyebrow image can located.
The final step after locating all FCPs is using equation (1) to calculate the Euclidean distance between certain FCPs to be the extracted geometric features. These lengths with the two mouth angles were chosen because they are the most lengths that convey facial expression information.

3.2.2 Appearance Features Extraction

The appearance features represent an important part of the facial expression feature. That is because of its holistic nature that deals with the whole face image. As mentioned before many approaches can be used to extract facial expression features such as edge orientation histograms that proposed in this paper. In this major step the normalized image (300X300 pixels) is reduced to (250X200 Pixels) by removing a 50 pixels from left, right and top to focus on the face without the hair. Then the "Canny" edge detection is applied to get an edge map after some image enhancements (histogram equalization and Gaussian filter). The edge map is divided into zones. The coarsely quantized edge directions are represented as local appearance features and more global appearance features are presented as histograms of local appearance (edge directions). The edge directions are quantized into 4 angular segments.
Finally, the face map is represented as feature vector of 64 components (4 histograms of 16 zones).

Figure1.

Block diagram of the facial expression recognition system



4. Conclusion

This paper presented an automatic system for facial expression recognition. It addressed the problems of how to detect a human face in static images and how to represent and recognize facial expressions presented in those faces. A hybrid approach is used for facial features extraction as a combination between geometric and appearance facial features. CNN based neural network is used for facial expression recognition. The proposed system can be tested for various databases, while testing it is known that 93.5% accurate in recognition for person dependent dataset. The future work will be the dealing with other types of images like image sequence and 3D images.

References

1. J. Bromley, I. Guyon, Y. LeCun, E. S¨ackinger, and R. Shah. Signature verification using a "siamese" time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7(4):669–688, 1993.

2. Cohn, J. F., Zlochower, A. J., Lien, J.J., & Kanade, T. (1998). Feature-Point Tracking by Optical Flow Discriminates subtle Difference in Facial Expression., *Proc. Int'l Conf. Automatic Face and Gesture Recognition, Nara, Japan, pp3396, April 14-16.*

3. Edwards, G.J., Cootes, T.F. & Taylor, C.J. (1998). Face recognition using active appearance models. *In proceedings of the European conference on computer vision.*

4. Gizatdinova, Y. & Surakka, V. (2006) .Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, no. 1, pp.135-139.*

5. Monwar,~~ M. & Rezaci, S~~ (2006). Pain recognition using artificial neural network, *Signal Processing and Information Technology, 2006 IEEE International Symposium on.*

6. A. Gepperth. Visual object classification by sparse convolutional neural networks. In Proceedings of the European Symposium on Artificial Neural Networks, 2006.

7. Ratliff, M. S., and Patterson, E. (2008). Emotion recognition using facial expressions with active appearance models. *International association of science and technology for development.*

8. S. Duffner and C. Garcia. An online backpropagation algorithm with validation error-based adaptive learning rate. In International Conference on Artificial Neural Networks (ICANN), volume 1, pages 249–258, Porto, Portugal, September 2007.

9. Lucey, P., Cohn, J., Lucey, S., Matthews, I. Sridharan, S. & Prkachin, K. M. (2009). Automatically detecting pain using facial actions, *IEEE.*

10. Ryan, A. Cohn, J. F., Lucey, S., Saragih, J. Lucey, P. Torre, F. & Rossi, A. (2009). Automated Facial Expression Recognition System, *IEEE*.