FACIAL EXPRESSION RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORK

Karthika
PG Scholar, Department of Computer Science &Engineering
Pandian Saraswathi Yadav Engineering College, Sivagangai, Tamil Nadu

K.R. Premlatha & D. Vanathi
Professor & Assistant Professor Department of Computer Science & Engineering
Pandian Saraswathi Yadav Engineering College, Sivagangai, Tamil Nadu

Abstract

Facial expression recognition is an important field of pattern recognition research. Traditional machine learning methods extract features manually. It has insufficient generalization ability and poor stability. Moreover, its accuracy is difficult to improve. In order to achieve better facial expression recognition, using deep convolutional neural network. To avoid over fitting, the network output uses a global average layer. Data enhancement on the dataset before training can improve the generalization ability of the model. Test the performance of network on the FER2013 emoticon dataset. Compared to other recognition algorithms, network has certain advantages. Finally, a facial expression recognition system is constructed by using the trained recognition model by using deep learning technique and SGD machine learning algorithm was implemented for optimization and to reduces the loss of the system and finally generate the result based on accuracy. The experimental results show that the system can effectively recognize facial expressions. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.

Keyword: *Machine Learning, Stochastic Gradient Descent, Conventional Neural Network*

Introduction

Facial expression is one of the primary non-verbal communication methods for expressing emotions and intentions. Ekman et al. identified six facial expressions (viz. anger, disgust, fear, happiness, sadness, and surprise) as basic emotional expressions that are universal among human beings. Automatizing the recognition of facial expressions has been a topic of study in the field of computer vision for several years. Automated Facial Expression Recognition (FER) has a wide range of applications such as human computer interaction, developmental psychology and data driven animation. Despite the efforts made in developing FER systems, most of the existing approaches either have poor recognition rates suitable for practical applications or lack generalization due to the variations and subtlety of facial expressions. The FER problem becomes even harder when we recognize expressions in videos. Numerous computer vision and machine learning algorithms have been proposed for automated facial expression recognition. It performs fine on classifying facial expressions collected in controlled environments and lab settings.

However, these traditional approaches mainly consider still images independently and do not consider the temporal relations of the consecutive frames in a video. In recent years, due to the increase in the availability of computational power, neural networks methods have become popular in the research community. In the field of FER, we can find many promising results obtained using Convolutional Neural Networks. While in the traditional approaches features were handcrafted, the CNNs have the ability to extract more appropriate features from the training images that yield in better visual pattern recognition systems. Therefore, it has been concluded that the CNNs are able to extract features that generalize well to unseen scenarios and samples.

Due to uniqueness of expressions for each person and insufficient number of examples in available databases, one shot methods have received attentions in recent years. In these methods, information about different categories are learned from one or very few samples. In this work, since inputs are sequences of frames, the number of training samples is considerably lower than frame based approaches. Although our method is not purely one-shot, it uses low number of samples in the training phase which makes it is highly generalizable and learns well even with few training samples. In this paper, we look at the problem of facial expression recognition as a two-step learning process. In the first step, we model the spatial relations within the images. In the second step, we model the temporal relations between consecutive frames in a video sequence and try to predict the labels of each frame while considering the labels of adjacent frames in the sequence. We apply our experiments on three facial expression datasets in order to recognize the six basic expressions along with a neutral state. All of these databases have few numbers of training samples especially for a CNN. Furthermore, we examine the ability of our model in cross database classification tasks. Residual connections introduced by He et al. have shown remarkable improvement in recognition rates. We propose a residual neural network for the first part of our network. For the second part, we use linear chain Conditional Random Fields (CRFs) model. Cascading of the aforementioned methods is a good approach in modeling facial expression. FACE detection and alignment are essential to many face applications, such as face recognition and facial expression analysis. However, the large visual variations of faces, such as occlusions, large pose variations and extreme lightings, impose great challenges for these tasks in real world applications. However, quite a few works indicate that this detector may degrade significantly in real-world applications with larger visual variations of human faces even with more advanced features and classifiers. Besides the cascade structure, introduce deformable part models (DPM) for face detection and achieve remarkable performance.

Existing Model

In the existing system, the module uses three different channels, increasing the diversity of extract features. Then, the feature maps of the three channel outputs are cascaded together. Finally, use this module to build Expression Net. Compare Expression Net to the classic Alex Net and other algorithms in the FER2013 dataset. The results show that the accuracy of Expression Net is significantly improved, and the parameters are greatly reduced. It proves that the module can effectively improve network performance and control model size. Using the FER-2013 dataset, a test accuracy of 64.6% is attained with this designed CNN model. The achieved results are satisfactory as the average accuracy on the FER-2013 dataset is 65% +/- 5%, and therefore, this CNN model is nearly accurate.

Proposed Model

The proposed model is introduced to overcome all the disadvantages that arise in the existing system. This system will increase the accuracy of the neural network results by classifying the fer2013 face expression image dataset using a deep learning algorithm. The data augmentation is used to generate the image samples for the train and test datasets. It enhances the performance of the overall classification results. Predict the facial expression and find the accuracy more reliable.

Objectives

To develop a facial expression recognition system based on convolutional neural networks (CNN) and Visual Geometry (VGG16) Group with data augmentation This approach enables us to classify seven basic emotions consisting of anger, disgust, fear, happiness, neutral, sad, and surprise from image data from the FER 2013 database. The facial expression recognition is realized through building and training the CNN network by Keras. To enhance the performance of the overall prediction results.

Image Acquisition and Resolution

The image acquisition procedure includes several issues, such as the properties and number of video cameras and digitizers; the size of the face image relative to total image dimensions; and ambient lighting. All of these factors may influence facial expression analysis. Images acquired in low light or at coarse resolution can provide less information about facial features. Similarly, when the face image size is small relative to the total image size, less information is available. The NTSC cameras record images at 30 frames per second. The implications of down-sampling from this rate are unknown. Many algorithms for optical flow assume that pixel displacement between adjacent frames is small. Unless they are tested at a range of sampling rates, the robustness to sampling rate and resolution cannot be assessed. Within an image sequence, changes in head position relative to the light source and variations in ambient lighting have potentially significant effects on face expression analysis. A light source above the subject’s head causes shadows to fall below the brows, which can obscure the eyes, especially for subjects with pronounced bone structure or hair. Methods that work well in studio lighting may perform poorly in more natural lighting (e.g., through an exterior window) when the angle of lighting changes across an image sequence. Most investigators use single-camera setups, which is problematic when a frontal orientation is not required. With image data from a single camera, out-of-plane rotation may be difficult to standardize. For large out-of-plane rotations, multiple cameras may be required.

Face Acquisition

With few exceptions, most AFEA research attempts to recognize facial expressions only from frontal-view or near frontal-view faces. Kleck and Mendolia first studied the decoding of profile versus full-face expressions of affect by using three perspectives (a frontal face, a 90° right profile, and a 90° left profile). Forty eight decoders viewed the expressions from 64 subjects in one of the three facial perspectives. They found that the frontal faces elicited higher intensity ratings than profile views for negative expressions. The opposite was found for positive expressions. Pantic and Rothkrantz used dual-view facial images (a full-face and a 90° right profile) which are acquired by two cameras mounted on the user's head. They did not compare the recognition results by using only the frontal view and the profile. So far, it is unclear how many expressions can be recognized by side-view or profile faces. Because the frontal-view face is not always available in real environments, the face acquisition methods should detect both frontal and non-frontal view faces in an arbitrary scene. To handle out-of-plane head motion, face can be obtained by face detection, 2D or 3D face tracking, or head pose detection. Non frontal view faces are warped or normalized to frontal view for expression analysis.

Facial Feature Extraction and Representation

Overview

After the face is obtained, the next step is to extract facial features. Two types of features can be extracted: geometric features and appearance features. Geometric features present the shape and locations of facial components (including mouth, eyes, brows, and nose). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face, such as wrinkles and furrows. The appearance features can be extracted on either the whole-face or specific regions in a face image.

Facial Expression Recognition

The last step of AFEA systems is to recognize facial expression based on the extracted features. Many classifiers have been applied to expression recognition such as neural network (NN), support vector machines (SVM), linear discriminant analysis (LDA), K-nearest neighbor, multinomial logistic ridge regression (MLR), hidden Markov models (HMM), tree augmented naive Bayes, Rank Boost, and others. Some systems use only a rule-based classification based on the definition of the facial actions. Here, we summarize the expression recognition methods to frame-based and sequence-based expression recognition methods. The frame-based recognition method uses only the current frame with or without a reference image (it is mainly a neutral face image) to recognize the expressions of the frame. The sequence-based recognition method uses the temporal information of the sequences to recognize the expressions for one or more frames. The recognition methods, recognition rates, recognition outputs, and the databases used in the most recent systems. For the systems that used more classifiers, the best performance for person independent test has been selected.

Conclusion

In this study, the fer2013 image dataset is used to find the facial emotion expression using a deep learning algorithm. In the first step, import the facial expression training and test dataset as input images and resize the image, then augment the image using a data image generator and convert the images into an array by dividing the default dimension using.255. Finally, the conventional neural network algorithm is applied to find the prediction result of facial expression, and the result is based on accuracy.

References

1. Jiaxing Lia, Dexiang Zhanga, Jingjing Zhanga, Jun Zhanga, Teng Lia, Yi Xiaa, Qing Yana and Dr. Mamoona Asghar, "Number Plate Recognition System, Facial Expression Recognition with Faster R-CNN ,The School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China.
2. Minh-An Quinn, Grant Sivesind, Guilherme Reis, Real-time Emotion Recognition from Facial Expressions CS 229 Stanford University.
3. Minchul Shin, Munsang Kim, and Dong-Soo Kwon, Baseline CNN structure analysis for facial
4. Expression recognition.