Recognizing and Expressing Affect

Project Report

By Xin Chen (ID 20705271)

SYDE 675

April 08, 2018

Recognizing and Expressing Affect

XIN CHEN

Abstract:

Keywords:

1 Introduction

In our daily life, affective competence with normal rational thinking and logical reasoning ability is playing an important role. At present the computer is still based on logical reasoning ability, if humans are able to bring the computer emotional ability, like bringing the computer a higher, comprehensive intelligence, establishing a harmonious environment for human-computer interaction, and giving the computer ability to adapt to human emotions, then to some extent, it will provide a lot of convenience to life. Shaping the artificial character and combining it with facial expression recognition, expression synthesis and artificial intelligence techniques can construct a virtual human proto-system which can continuously express emotions.

Affective computing refers to the calculation which arises from or deliberately influences emotions, and its main purpose is that computers are able to acquire the ability of recognizing, understanding and expressing affect. Affective computing will receive facial expression and signal from the change of people's bodies by various sensor and then recognizing these signals by affective models, which is for understanding human emotions and make an appropriate response.

This project will focus on emotion modeling, emotion recognition and its related applications. Emotional models are created to make computers smarter, more friendly, and more capable. Therefore, we use a variety of methods to establish emotional models, which is for bringing humans and computer better interaction. In addition, collecting some relevant information like reading papers, journals and magazines, can make us understand the current development of the technology and future applications.

2 Research status of facial expression recognition

Facial expression recognition is the first step of affect recognition, which plays an important role of results of affect recognition. Nowadays, most of researches are under the circumstance of having clear and discernible faces.

There are two ways for us to locate and detect a face in a picture. One, we can regard a face as a whole to recognize. Two, we can detect some important characters of faces to recognize them.

Pantic and Rothkrantz viewed a face as a whole to recognize. They focused on the shapes and colors of faces. However, there are many disadvantages like they are not able to eliminate the interference of glasses or hairs on faces.

In 2001, Paul Viola held the view that people could apply AdaBoost algorithm to recognizing faces. AdaBoost algorithm would select some vital characters to construct classifiers and rapidly focused on some areas which are like faces. This method not only improved the speed of detection but also increased the detection rate. In addition, most of facial recognition systems are based on AdaBoost algorithm and then improve themselves.

2.2 feature extraction

Nowadays, the technology of facial recognition has improved a lot and it also meets needs of researches, so the researches of facial affect recognition now are mostly focus on how to make sure different emotions match with the different correct facial expression characters. In general, facial expressions are classified as surprise, fear, abhor, anger, happy and sad. Therefore, most of researches are focusing on facial feature extraction.

The feature extraction of facial expressions is important, and the feature directly determines the speed and efficiency of recognition. For static images, the deformation features of expressions need to be extracted. Such methods generally only consider the spatial information and the face geometry information of a single frame image, and the calculation will be simple. The image sequence is different. It is necessary not only to consider the characteristics of deformation in each frame, but also to consider the movement characteristics of the image. This method has a high recognition rate, but it is more difficult to calculate.

Considering the different ranges of images, features can be further divided into two types, global and local. The global feature extraction method directly extracts the motion features and deformation features of the entire face. Based on the local feature extraction method, the face is divided into several sub-regions. Each sub-region is extracted and then the results are concatenated. The facial deformation feature extraction method refers to the neutral expression face, and the deformation information is obtained by comparing the neutral face with the current face. The combination of information should be reflected in two aspects: facial shape changes and facial lines change. This type of extraction can be roughly divided into three categories:

(1) Method based on geometric features. The human face shows different expressions through changes in the eyes, nose, mouth, etc. It is possible to create a geometric model of a specific area of the face and extract the geometric features of the expression. Pantic M. et al. used multiple detectors to spatially locate human faces and facial organs, and finally extracted ten human face fiducials and 19 human facial organ fiducials to identify 32 muscle motor units.

(2) Based on Gabor wavelet method. Wavelet transform essentially uses a set of filters of different scales to filter the signal and then decompose the result into different frequency bands for processing and analysis. The effect of this method is similar to the multi-channel filtering model in the human visual system. Therefore, in the face expression recognition application, wavelet transform is used to extract the wavelet feature of the image. The Gabor wavelet is insensitive to changes in global brightness and affine transformations. It reflects image structure characteristics. The Gabor wavelet coefficient method with uniform sampling points and the Gabor wavelet transform-based recognition method have achieved good results in facial expression recognition. The Gabor wavelet method is less affected by the change of illumination, it can detect multi-directional and multi-scale texture changes, and it is widely used in feature extraction of face expressions. Donato compares several methods to identify the performance of facial AU, and the results show that wavelet transform is superior to other methods. Yao Wei has proposed a two-layer Gabor feature selection method. Firstly, the high-dimensional vector is filtered by improving the variance ratio as the distinguishing ability of the evaluation feature, and then the features of the

obtained feature subset are selected by AdaBoost. Finally, the most discriminative feature is selected to achieve the purpose of reducing the feature dimension.

(3) Model-based approach. A model is used to describe the structure of a human face. A geometric model (point model) is a simple model method. Tim Cootes, a scientist at the Department of Image Science-Biology Engineering at the University of Manchester, UK, first proposed the AAM method, which is also a widely used face feature extraction method. AAM differs from the Active Shape Model (ASM) because it takes into account the surrounding area covered by the object shape, in addition to the information on the edge of the object shape. The AAM method statistically models the textures and shapes of interest objects so that the resulting model contains as many valid objects as possible. The AAM method consists of two parts: building an object model and searching for new objects. Firstly manually locate the feature point object in the image. After the shape of the object is formed, the texture of the object is generated based on the image covered by the shape. A statistical model is created for all the textures and shapes of the objects to be learned, and the Appearance Model is created using the Texture Model and the Shape Model. Create new objects by changing the appearance model's parameters. When the texture of the object to be detected and the object error generated by the model are the minimum, the position and shape information of the new object are obtained, and the parameters of the model are assigned to the new object. The model-based method can obtain more reliable face features, but it has the disadvantages of difficulty in obtaining initial points and large calculation volume. Zuo Kunlong analyzed the feasibility of facial expression recognition (FER) using facial expression features extracted from M. He tried to use the FER based on this feature vector to locate the eye area based on the characteristics of the human face image, and then use AAM's optimization algorithm to obtain the new object's characteristics. This not only improves the accuracy of positioning, but also greatly reduces the time required to locate new objects from the M method.

(4) Subspace-based methods. The dimensions of facial expression images are generally very high. At the same time, the distribution in high-dimensional space is very scattered, which is not conducive to classification, and the computational complexity is also quite large. In order to obtain a more compact distribution of

the image, it is necessary to reduce the dimension of the original high-dimensional space, and finally perform classification and recognition of the expression in the low-dimensional subspace. The idea of subspace analysis is to reduce the compression of the original data to a low-dimensional space, to make the distribution of data in the subspace more compact, and to greatly reduce the subsequent calculation.

Subspace-based facial expression recognition methods can be divided into two kinds of nonlinear and linear space transformation. The linear subspace methods that have been successfully applied in facial expression recognition include: Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Independent Element Analysis (CA). Nuclear-based nonlinear subspace analysis methods include: Kernel Principal Component Analysis (KPCA) and Nuclear Fisher Discriminant Analysis (KFDA).

(5) Based on the local binary mode method. The Local Binary Patterns (LBP) can describe the texture features of the graph well. It obtains local texture features by comparing the gray value of any point in the image with the surrounding points. Due to its good characteristics of anti-rotation and brightness changes, it has been widely used in many fields of pattern recognition in recent years. In terms of precise positioning of faces, face detection, image content recognition, texture recognition, and face recognition, good results have been achieved. Compared with the Gabor wavelet feature, extracting LBP features in face images will be faster, and the dimension will be much smaller, and at the same time, the face information will be effectively preserved. Shan uses LBP to extract the texture features of the face image. Firstly she calculated the LBP histograms of the face region in blocks, and then uses these LBP histograms to string together all of them as facial feature results, which has a better recognition effect. Sun Ning et al. proposed a local feature extraction method based on 2-D PLS (2DPLS) and applied it effectively in facial expression recognition. The method first used the LBP operator to extract the texture features of all the sub-regions of an image, and then combined them into a local texture feature matrix. 2DPLS adapted the corresponding relation matrix so that the matrix form can adapt to the sample, and at the same time, it showed the importance of different local information in the face. Cui Jie and others used LBP to extract facial features and perform coarse-to-

fine facial expression classification. At the rough classification stage, two expressions were selected as preliminary classification results (candidate expressions). Then in the fine classification stage, the final classification was determined by calculating the weighted chi-squared value.

The facial motion feature extraction rule was to compare the changes of face information in a period of time. The extraction of such features includes feature point tracking, optical flow, differential images, and motion models. The optical flow method is an earlier and more widely used method. It uses the optical flow to express the characteristic movement of different expressions, and uses the optical flow of the characteristic point as the expression feature. For example, Lien et al. used a wavelet-based multi-resolution dense optical flow to analyze the whole face motion. However, the optical flow method has a large amount of calculations and is sensitive to the discontinuity of motion. It may take a lot of time to calculate the optical flow in the entire face area. Feature point tracking is a traditional algorithm for extracting facial features. It is different from the optical flow method and does not track every pixel point in the image. It only tracks certain interest points. The marker point method is similar to the feature point tracking method, but more is used in areas where the cheeks do not have much texture information.

2.3 Expression recognition

Expression classification is the final stage of expression recognition. It is closely related to the extraction methods of facial expression features. The purpose is to classify different expressions into the corresponding expression classes. Some scholars have suggested that how to select and design classifiers depends on whether they use time information. The classification method using time information is called time domain airspace method, and the classification method that does not use time information is called airspace method, similar to the static and dynamic method of division. Typical of the airspace method is the artificial neural network. The neurons in the input layer correspond to the input face data, and the output neurons correspond to expression classes. However, neural network training is difficult, so Kaiser proposed a rule-based neural network. Because the airspace method only uses eigenvectors, the general classification method can be classified as this type. Wang et al. used the adaboost algorithm to construct a classifier, and Buciu et al. used a support vector machine (SVM) as a classifier, and Guo et al. used a linear programming method to construct a classifier. At the same time, the expression is also a dynamic process, and the higher recognition rate can be obtained by considering the use of dynamic information when designing and selecting methods. The time domain airspace method effectively uses dynamic information. Currently Hidde Markov Model (HMM) is often used in facial expression recognition to better describe the dynamic sequence. In addition, recursive neural networks also have many successful applications in facial expression recognition systems.

2.4 Emotion research and application

Emotion computing is a new multidisciplinary research area. It includes computer science, sensor technology, psychology, physiology, cognitive science, philosophy, and so on. The ultimate goal of emotional computing is to make computers feel like humans. There are many problems that need to be solved to achieve this goal now. The key technologies studied include human emotion recognition, emotional modeling, and effective expression of emotional perception. Emotional experience includes both psychological and physiological processes. Emotional research attempts to make an overall explanation of the psychological and physiological processes of emotions and the relationships between them. Psychologists put forward a variety of perspectives from different perspectives and research perspectives, thus forming many emotional theories. On the basis of emotional theory, people have further established an emotional model that can explain the process of emotional development. These theories include physiological response theory, stimulus-response theory, motivation theory, facial expression theory, and subjective evaluation theory. At present, the cognitive evaluation theory of emotion is most accepted by people. The construction of emotional subjects is mostly based on certain cognitive evaluation theories, and artificial intelligence scholars have also pay more attention to it. According to the theory of cognitive evaluation, when the subject evaluates an event that is subjectively important, it will produce an emotional experience, which in turn will generate emotions. This evaluation process is more subjective, and it depends on the beliefs of the subject, specific goals and norms. Different subjects have different internal psychological structures. As a result, even the same external stimulus, the interpretation may be different. The ultimate emotions will be produced through their own cognition and subjective evaluation of the stimuli.

The most influential of these evaluation theories is the OCC theory proposed by Ortony, Collins and Clore. The OOC model is the first model designed for computer implementation. They assume that emotions result from a process called cognitive evaluation. Evaluation depends on three components: events, objects, and subjects. According to the subjective goal, the events in the objective world are rated as satisfactory or not. Subjects are rated like or dislike based on the subject' s attitude. The behavior of the subject itself or other subject is rated as favored or disapproved according to a set of criteria. These evaluation variables eventually form a hierarchy of 22 emotions.

The synthesis of happy emotions in the OCC model is expressed as follows:

Let D(p,e,t) represent the degree of expectation that person p wants event e at time t. If event e is good, it will return a positive value.

Let $P_j(p,e,t)$ express the possibility of happiness, and $I_g(p,e,t)$ is the total combination of intensity variables (eg, expectation, approximation, reality). An example of a rule that produces happy is:

If, $D(p,e,t)>0$

Then, $I_j(p,e,t)=F_j(D(p,e,t), I_g(p,e,t))$

And $F_j()$ is specified as a function that represents happiness.

Other emotions such as sad can also be implemented with similar rules. $P_d()$ can be changed by changing the expectation of negative conditions in D(p,e,t), $I_j(p,e,t)= F_d(D(p,e,t),Ig(p,e,t))$.

The above formula does not cause state happiness, but it can trigger another rule, thus changing the intensity of happiness I_j. After a threshold is given:

If, $P_j(p, e, t)>T_j(p, t)$

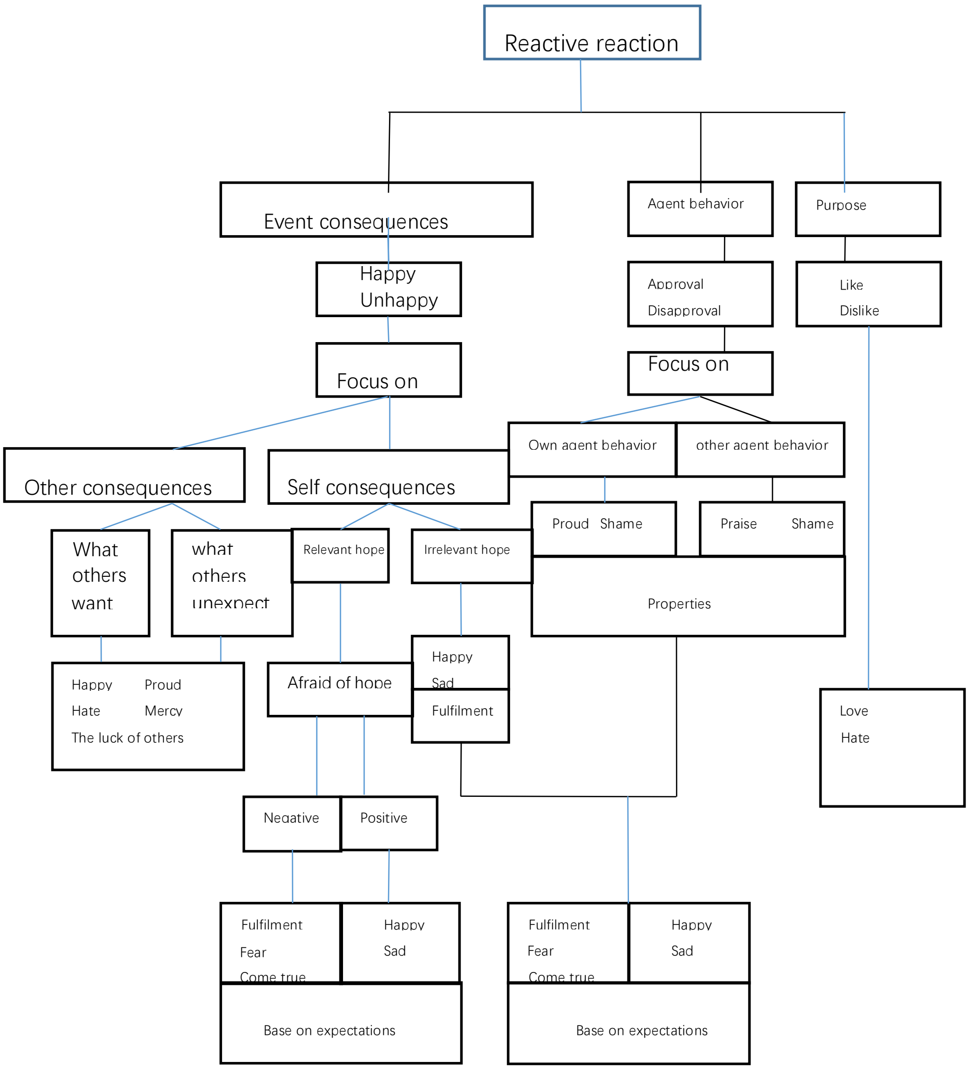
Then, $I_j(p, e, t)=P_j(p, e, t)—T_j(p, t)$

Plus, $I_j(p, e, t)=0$

This rule activates happy emotions. When intensity exceeds a happy threshold, the resulting intensity can be mapped into happy emotions.

Happiness is the simplest example. The realization of other emotions in the OCC model is more complicated.

OCC structure as shown:



OCC model emotional cognitive structure

---models（markov， hmm）

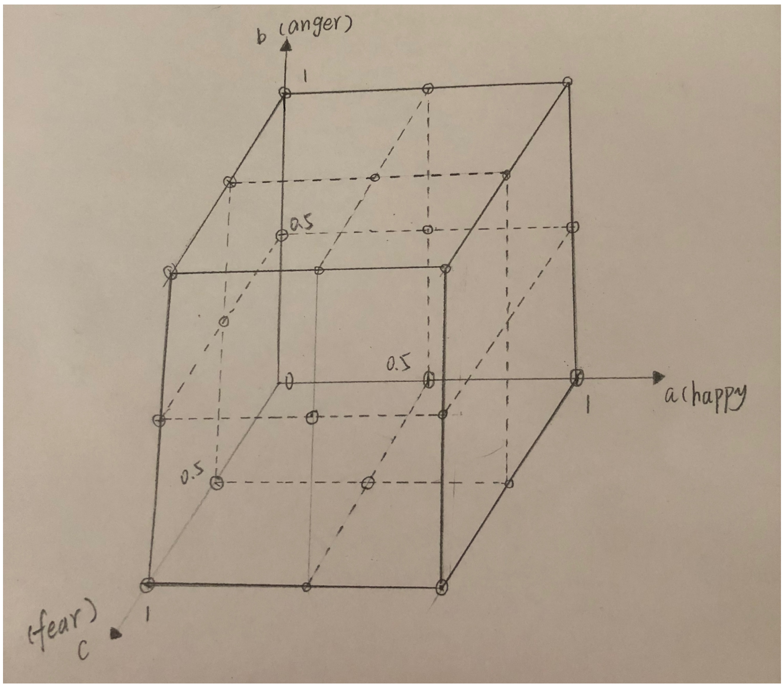
This paper analyzes the nature of the facial expressions and gives a qualitative descriptio n of the corresponding facial expression space, and then proposes a new facial expressi on space model with the characters of both discrete affective space model and continuou s affective space model.

We will focus on discrete affective space model.

Firstly, we will construct a three dimensional model based on fear, anger and happy, and any emotional status will match a certain point on the 3 dimensional space. In order to realize it, we will have to simplify that by discretizing. We can set that every basic emotion only have three intensity, which means that every dimensional will have 0, 0.5, and 1, three different values. Take happy for example, there are three strength, unhappy, a little happy and happy. Therefore, there are 27 discrete emotional status. We can define happy as a, anger as b, fear as c, so, there are a$\epsilon(0, 0.5, 1)$, b $\epsilon(0, 0.5, 1)$, c $\epsilon(0, 0.5, 1)$.

No.	Status of Emotion
1	Quiet(0,0,0)
2	A little happy(0.5, 0, 0)
3	A little angry(0, 0.5, 0)
4	A little fear(0, 0, 0.5)
5	A little happy, a little angry(0.5, 0.5, 0)
6	A little angry, a little fear(0, 0.5, 0.5)
7	A little happy, a little fear(0.5, 0, 0.5)
8	A little happy, a little fear, a little angry(0.5, 0.5, 0.5)
9	Happy(1, 0, 0)
10	Angry(0, 1, 0)
11	Fear(0, 0, 1)
12	
13	
14	

15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	Happy, fear, angry(1, 1, 1)



During the 27 points, origin of coordinates means quiet or no emotions. Some vertex also means quiet or some other emotion status. It is clear that this model is not completed. When someone's emotion go through this space, it is easy for us to find its statistical natures. This process we also call it Markov. Therefore we can use Markov model to describe how emotions change.

There is probability of changing emotion status in Markov model, so in this emotion model, there are 27 $P_{i,j}$(i,j $\in [1,2,3 \dots \dots ,27]$) and they construct 27 dimensional Probability matrix A_p.

$$A_p = \begin{bmatrix} P_{1,1} & \cdots & P_{1,27} \\ \vdots & \ddots & \vdots \\ P_{27,1} & \cdots & P_{27,27} \end{bmatrix}$$

$P_{i,j}$(i,j $\in [1,2,3 \dots \dots ,27]$) is the probability of i_{th} status to j_{th} status. Additionally, there is a relationship among them:

$$\sum_{i=1}^{27} p_{i,j} = 1, i \in [1, 2, 3 \dots \dots , 27]$$

We can conclude that if there are m emotions and there will be m dimensional emotional space, and for every emotion there are n levels, which means that there will be n^m emotion status. Reagard l=n^m, we can find that:

$$A_p = \begin{bmatrix} P_{1,1} & \cdots & P_{1,l} \\ \vdots & \ddots & \vdots \\ P_{l,1} & \cdots & P_{l,l} \end{bmatrix} \text{ and } \sum_{i=1}^{l} p_{i,j} = 1, i \in [1, 2, 3 \dots \dots , l]$$

In Markov model, the probability of change emotion will be influenced by many factors like personal characters, conscious stimulation. For example, under the positive simulation, the probability of changing a certain emotion to positive emotion will be larger than that of changing a certain emotion to negative emotion.

d-s theory

The D-S theory is a generalization of Bayesian inference, mainly by using the Bayesian conditional probability in the probability theory, and it needs to know the prior probability. While D-S evidence theory does not need to know a priori probability, and it can express "uncertainty" well, so it is widely used to deal with uncertain data.

Combine emotional space model with d-s evidence theory. Someone's emotion will always stay at a certain point on this space and it will probably change to another point. But this process not only relate to its original position but also simulation from outer space. For example, it will be influenced by sounds, weather, pictures, smell and so on. Firstly, this model will capture outer simulation by sensor, and then D-S theory will be applied into it for combining outside emotion information, which finally will promote the transfer of emotion status and reach a new state of emotion.

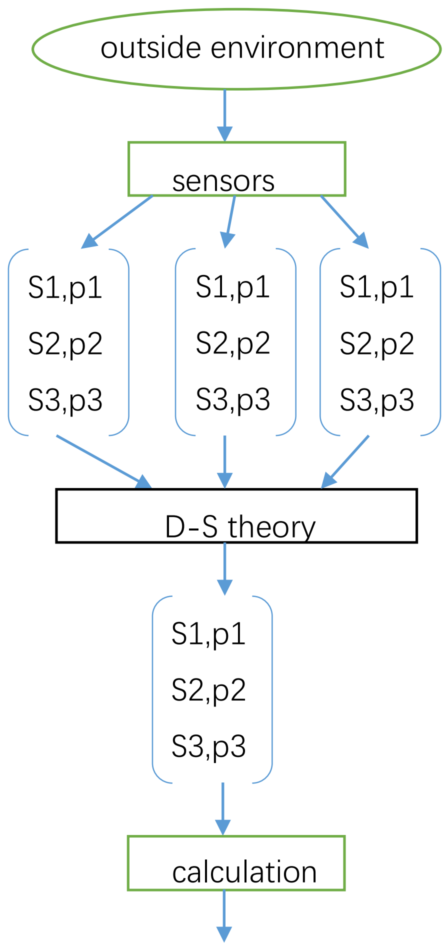
---facial

Combine emotional space model with d-s evidence theory.

Firstly, this model will capture outer simulation by sensor, and then D-S theory will be applied into it for combining outside emotion information, which finally will promote the transfer of emotion status and reach a new state of emotion.

S = the strength of simulation from outside environment

P = basic credibility



Picture 2 Emotional Information Processing Flow

Firstly, we should set an initial state, picking a start point in the three-dimensional space.

4 using CNN convolutional neural network to recognize facial expression

I will design a specific experiment on how to recognize facial expression

First of all, I download some dataset from the website, and some of them are as training data and others are testing data. (For convenience, we used 3000 pictures as training data and 5000 pictures as testing dataset)

There are five steps for us to do:

1. we can read the data downloaded from website.
2. We will get and save the labels and pictures, setting up the batch for providing data when training network.
3. Construct CNN convolutional neural network.
4. Train it.
5. Check it and calculate the accuracy rates.

----sound

---body

2 The current current status of facial expression recognition

Humans are able to communicate with each other by observing their facial expression. It is mainly because people have already known some main features of facial expression and they are able to tell a person if he or she is happy or not. Pantic M and Rothkrantz L[1] hold the view that

Reference

[4] Pantic M，Rothkrantz L. Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence，2000，22(12)：1424-1445.