Affective Computing: A Review

Conference Paper · October 2005

DOI: 10.1007/11573548_125 · Source: DBLP

2 authors, including:

Jianhua Tao
Chinese Academy of Sciences
320 PUBLICATIONS **2,278** CITATIONS

Affective Computing: A Review

Jianhua Tao and Tieniu Tan

National Laboratory of Pattern Recognition (NLPR), Institute of Automation,
Chinese Academy of Sciences, P.O.X. 2728, Beijing 100080
{jhtao, tnt}@nlpr.ia.ac.cn

Abstract. Affective computing is currently one of the most active research topics, furthermore, having increasingly intensive attention. This strong interest is driven by a wide spectrum of promising applications in many areas such as virtual reality, smart surveillance, perceptual interface, etc. Affective computing concerns multidisciplinary knowledge background such as psychology, cognitive, physiology and computer sciences. The paper is emphasized on the several issues involved implicitly in the whole interactive feedback loop. Various methods for each issue are discussed in order to examine the state of the art. Finally, some research challenges and future directions are also discussed.

1 Introduction

Affective computing is trying to assign computers the human-like capabilities of observation, interpretation and generation of affect features. It is an important topic for the harmonious human-computer interaction, by increasing the quality of human-computer communication and improving the intelligence of the computer.

The research on affect or emotion can be traced from nowadays to 19 century [2]. Traditionally, "affect" was seldom linked to lifeless machines, and was normally studied by psychologists. It is quite new in the recent years that the affect features were captured and processed by the computer. The affective computing builds an "affect model" based on the various sensors-captured information, and builds a personalized computing system with the capability of perception, interpretation to human's feeling as well as giving us intelligent, sensitive and friendly responses.

To get the impression of state of art of the research in affective computing, the paper briefly summaries some key technologies for the research during last several years, such as emotional speech processing, facial expression, body gesture and movement, multimodal system, affect understanding and generating, etc. A brief discussion will also be made in each topic. Furthermore, the paper also introduces some related projects in the world, which gives the clearer impression on the current/past research work and applications. Based on above summary and analysis, the paper discussed some hot research topics which might be a big challenge to improve the current research work.

The paper is organized as follows. Section 2 describes the recent development of the related key technologies. Section 3 describes some related projects in this area. Section 4 summaries the hot research topics. The final conclusion of the paper will be made in section 5.

J. Tao, T. Tan, and R.W. Picard (Eds.): ACII 2005, LNCS 3784, pp. 981 – 995, 2005.

2 State of Art of Key Technologies

The standard procedure of affective interaction consists of affect information capture and modeling, affect understanding and expression etc. As we know, people express the affects through a series of action on facial expression, body movements, various gestures, voice behavior, and other physiological signals, such as heart rate and sweat, etc. The following parts will try to review the most active key technologies in these area, such emotional speech processing, facial expression recognition and generating, body gesture and movement, multimodal system, affect understanding and generating, etc.

2.1 Emotional Speech Processing

For emotional speech processing, it is a widely known fact that the emotional speech differs with respect to the acoustic features[43]. Some prosody features, such as pitch variables (F0 level, range, contour and jitter), speaking rate have been analyzed by some researchers [46]. Parameters describing laryngal processes on voice quality were also taken into account in someone's work [42]. Tato [48] made some experiments which showed how "quality features" are used in addition to "prosody features.

The above acoustic features are widely used for the research of emotion recognition with the pattern recognition methods. For instance, Dellaert [45] used prosody features and compared three classifiers: the maximum likelihood Bayes classification, kernel regression, and k-nearest neighbor in emotion recognition for sadness, anger, happiness and fear. Petrushin [46] used vocal parameters and a computer agent for emotion recognition. Lee [47] used liear discriminant classification with Gaussian class-conditional probability distribution and k-nearest neighborhood methods to classify utterances into two basic emotion states, negative and non-negative. Yu [49] used SVMs for emotion detection. A average accuracy of 73% was reported. Nick [40] proposed a perception model of affective speech utterances and shown that there are consistent differences in the acoustic features of same-word utterances that are perceived as having different discourse effects or displaying different affective states. The work proposed that rather than selecting one label to describe each utterance, a vector of activations across a range of features may be more appropriate.

For emotion generating with speech synthesis, Mozziconacci [7] added emotion control parameters on the basis of tune methods resulting in higher performance power of voice composing. Cahn [8], by means of a visualized acoustic parameters editor, achieved the output of emotional speech with manual inferences. Recently, some efforts have been down with the idea of the large corpus. A typical system was finished by Campbell [50], who created an expressive speech synthesis with a five years' large corpus and gave us an impressive synthesis results. Schroeder[51], Eide[44] generated a expressive TTS engine which can be directed, via an extended SSML, to use a variety of expressive styles with about ten hours of "neutral" sentences. Optionally, rules translating certain expressive elements to ToBI markup are manually derived. Chuang[52] and Tao[11] used emotional keywords and emotion trigger words to generate the emotional TTS system. The final emotion state is determined based on the emotion outputs from textual content module.

The results were especially used in the dialogue systems to improve the naturalness/expressiveness of the answering speech.

Till now, most of the researches on emotional speech are still focused on some typical acoustic features analysis in different languages. Some work in emotion classification systems and rule based emotional speech synthesis systems have been done [18], however, the lack on the capture and the analysis of more detailed/reliable physiological features limits the further improvement of the research. The people express the feeling not only by the acoustic features, but also with the content they want to say. Different words, phrases and syntactic structures, etc. can make lots kinds of expression results and styles. Though, some language cognition has been done by some psychologist before [29], lots of work is still needed for the integration of these two research topics.

2.2 Facial Expression

Facial expressions and movements such as a smile or a nod are used either to fulfill a semantic function, to communicate emotions, or as conversational cues. Similar as speech processing, the research of facial expression consists of works on coding, recognition and generation, and have been done for a very long history. For instance, Etcoff [9] parameterized the structure of the chief parts of human's face through 37 lines, which enables people to roughly tell the affect status of faces; Ekman [39] built facial action coding system. They classified human's facial expressions into many action units. With this method, he described facial expressions with six basic emotions, joy, anger, surprise, disgust, fear and sadness. Currently, most of the facial features can be found from the definition of MPEG-4. MPEG-4 allows the user to configure and build systems for many applications by allowing flexibility in the system configurations, by providing various levels of interactivity with audio-visual content [29][30]. In this standard, both mesh model [33] or muscle model are used to create 3-D facial models .

To do the facial expression analysis, most of the facial features were captured by the optical flow or active appearance model. Lyons [54] applied the supervised Fisher linear discriminant analysis(FDA) to learn an adequate linear subspace from class-specified training samples and the samples projected to this subspace can be best separated. Principal component analysis (PCA) [20] and independent component analysis (ICA) [25] have been used to for the expression classification. For facial expression recognition, there are many other methods, such as, Gabor wavelets [54], neural network [53], Hidden Markov Models (HMM) [30], Point Distribute Model (PDM), optical flow, geometrical tracking method, EGM method and so on. Among them, the Gabor representation has been favored by many researchers due to its good performance and not sensitive to the face posture and the lighting background. [54].

The pioneering work on facial animation was done by Frederic I. Parke in the 1970s. In the last decade the quality of facial animations has improved remarkably due to the development of hardware and corresponding software. But in these days, the generating of lifelike animated faces still remains an open issue. Many researchers used methods based on images [34][36], visemes[37], FAPs[43], PCs[31][32], 3D coordinates[37], 3D distance measurements [30][31] or optical flows[38] to generate facial expression. Normally, the face expression should be synchronized with speech when people express their ideas or feeling, it shows the audio-visual mapping is the

key component to generate a vivid talking head system. There are mainly two approaches for the synthesis: via speech recognition or driven by speech directly. The first approach divides speech signal into language units such as phonemes, syllables, words, then maps the units directly to the lip shapes and concatenates them. Yamamoto E.[30] recognized phonemes through training HMM, and mapped them directly to corresponding lip shapes, through smoothing algorithm, the lip movement sequence is obtained. The second approach analyzes the bimodal data through statistical learning model, and finds out the mapping between the continuous acoustic features and facial control parameters, so as to drive the face animation directly by a novel speech. Massaro [29] trained ANN to learn the mapping from LPCs to face animation parameters, they used current frame, 5 backward and 5 forward time step as the input to model the context. Many other methods have also been tried, such as, TDNN[29], MLPs[32], KNN[31], HMM[30], GMM, VQ[30], Rule-based [37][38].

Recently, a new approach for audio-visual mapping has arisen [36], which is inspired from speech synthesis [35]. This method means to construct new data stream by concatenating stored data units in training database. It has advantage of that the synthesis result appears very natural and realistic. But even for that, the lip movement is still the focus in most of the research. Full facial expression, especially the correlation between facial expression and more acoustic features, such as prosody, timbre, etc. has seldom been touched. It needs more work in processing the features which we ignored before.

2.3 Body Gesture and Movement

Body gesture and movement is defined by the positions of body arthroses and their changes with time. Currently, the work for gesture processing is more focused on the hand tracking. Hand gestures can convery various and diverse meanings, to enhance the mood or to behave as a sign language. Traditionally, there are two methods, apparentness methods [15] and 3-D modeling methods [17]. The apparentness based method makes out the model by analyzing apparent features of hand gestures from 2-D images, while the 3-D methods do the tracking in real 3D environment. Compared to 3-D methods, the apparentness methods are less complicated, and more easy to be used in real-time computation, however more efforts should be done to adapt the method into high noise background and the real application. Some efforts have been done to adopt mixed modeling methods and describe the features of static hand gesture with multiple features (such as local profile features and overall image matrix features) [16]. It shows the higher and more robust tracking results.

In order to realize body gesture and movement based from image sequences, the key point is how to confirm the positions of body arthroses according to given image information. The existing methods normally require some limitation [17], such as, (a) different dress colours according to different body arthroses; (b) simple moving directions; (c) simple backgrounds; (d) some manual initial markings. With these methods, the profile of the target body is picked up at first, and then virtual framework that is similar to real body framework is taken out through energy function. After that, arthroses positions are determined based on the virtual framework by using anthropotomy knowledge. The energy function can restrain some background noises, and has low requirement for the preciseness of the fetched body profile. In addition, some people enable the computers to more accurately capture

data of face and body's rapid movement by some auxiliary equipment like electromagnetic inductor [14] and optical reflection signs [13]. Till now, the work is still a difficult subject in computer vision's area, especially in real application. Concerning the capture of body gesture and movement, in addition to further improvement of the capture accuracy and efficiency of parameters, how to obtain more robust and subtle body-language is still an urgent difficult problem for affective computing.

2.4 Multimodal System

As we can imagine, the direct human to human interaction is, by definition, multimodal interaction in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours. Multimodal systems are convinced by most of the researchers to improve the results of affect recognition/understand and to generate more vivid expressions in human computer interaction [10][12]. Multimodal systems are able to meet the stringent performance requirements imposed by various applications [10]. Such as, in biometrics recognition systems, Brunelli et al. [55] describe a multimodal biometric system that uses the face and voice traits of an individual for identification. Their system combines the matching scores of five different matchers operating on the voice and face features, to generate a single matching score that is used for identification. Bigun et al. [56]develop a statistical framework based on Bayesian statistics to integrate information presented by the speech (text-dependent) and face data of a user. Kumar et al. [57] combined hand geometry and palmprint biometrics in a verification system. A commercial product called BioID [58] uses voice, lip motion and face features of a user to verify identity. Jain and Ross [59] improved the performance of a multimodal system by learning user-specific parameters. General strategies for combining multiple classifiers have been suggested in [60] and [61]. There is a large amount of literature available on the various combination strategies for fusing multiple modalities using the matching scores (see for example [62]).

In human computer interaction applications, it has been widely used for smart room, virtual reality, etc. Among them, the ubiquitous computing [65] might be the most representative application of this technology, which encompasses a wide range of research topics, including distributed computing, mobile computing, sensor networks, human-computer interaction, and artificial intelligence.

The multimodal technology is just arisen in recent years, most existing system are lack of efficient method to integrate the different channels, the synchronized control modeling for multi-channel information processing are still not well solved. More work should be done in the parameters integrations.

2.5 Affect Understanding and Cognition

The affective understanding module is the next in logical sequence after the recognition module. The affective understanding may contain the functions by absorbing information, remembering the information, modeling the user's current mood, modeling the user's emotional life, applying the user affect model, updating the user affect model, building and maintaining a user-editable taxonomy of user preferences, featuring two-way communication with the system's recognition module,

eventually building and maintaining a more complete model of the user's behavior, eventually modeling the user's context, providing a basis for the generation of synthetic system affect, ensure confidentiality and security. [63]

The OCC model [41] might be one of the most successful models from all of the work. It classifies the people's emotions as a result of events, objects and other agents under three categories. People are happy or unhappy with an event, like or dislike an object, approve or disapprove an agent. There are 22 detailed emotions under the three emotion categories. Though the OCC model provides three groups of emotions depending on reactions to external things, it is really hard to do that in real environments, where we get more complicated reactions. Sometimes, some reactions may involve all of the emotions in the three categories. For example, when people see their neighbors beat the child, they may feel distressed and do not wish such an event (beating the child) to happen, and feel it a pity that the child is beaten. They may also reproach their neighbors for violating the human rights and contempt them for beating their child. Finally, they may start to hate their neighbor because of the event. From this process, we can see that people often experience more than one emotion because of complex external environments, instead of just one single emotion state. Therefore, it is difficult for us to understand the emotional experience of sadness and happiness mixed together and surprised happiness.

Affects are closely related to cognition. Psychologists have always been exploring this issue for a long time. In recent years the researchers from computer sciences also hope to verify the relations between affects and cognition through various experiments, for instance, the emotion group of Geneva university designed a set of computer games dealing with questions and answers. Experiment participants experience emotional changes as playing the games. Their facial expressions and sounds emitted during the game are collected as samples to be analyzed. This kind of games triggers emotions that can be used to help researchers study and explore the interaction of emotion triggering and cognition levels. Similar experiments were also conducted by UIUC in their multimodal interaction system.

Though the experiment designs and small samples are preliminary, the experiment shows that the close relation between emotions and cognition are being more valued by emotion researchers. From the preliminary theoretical framework to the current preliminary experiment, human beings are gradually revealing the secrets of their complex brains. With this progress, we are able to go deep into our brains, better control emotions triggered by brains, avoid the damaging behaviors brought about by passive and negative emotions, help mentally and psychologically ill patients overcome emotional shadows and make our psychological world more beautiful.

3 Projects

Although the affective computing is a new concept in recent years, there are already some related projects. We cannot list all of them, but only summarizing some of them according to author's experiences. They are described in the following.

3.1 HUMAINE (EU Project)

HUMAINE (Human-Machine Interaction Network on Emotion) is a Network of Excellence in the EU's Sixth Framework Programme. The project aims to lay the

foundations for European development of systems that can register, model and/or influence human emotional and emotion-related states and processes - 'emotion-oriented systems'. Such systems may be central to future interfaces, but their conceptual underpinnings are not sufficiently advanced to be sure of their real potential or the best way to develop them. One of the reasons is that relevant knowledge is dispersed across many disciplines. HUMAINE brings together lots of experts from the key disciplines in a programme designed to achieve intellectual integration. It identifies six thematic areas that cut across traditional groupings and offer a framework for an appropriate division - theory of emotion; signal/sign interfaces; the structure of emotionally coloured interactions; emotion in cognition and action; emotion in communication and persuasion; and usability of emotion-oriented systems. [21]

3.2 Affective-Cognitive Framework for Learning and Decision-Making (MIT Affective Computing Research Group)

The project aims to redress many of the classic problems, that most machine learning and decision-making models, however, are based on old, purely cognitive models, and are slow, brittle, and awkward to adapt, by developing new models that integrate affect with cognition. Ultimately, such improvements will allow machines to make smart and more human-like decisions for better human-machine interactions. [22]

3.3 Oz Project (CMU)

Oz is a computer system that allows authors to create and present interactive dramas. The architecture of the project includes a simulated physical world, several characters, an interactor, a theory of presentation, and a drama manager. A model of each character's body and of the interactor's body are in the physical world. Outside the physical world, a model of mind controls each character's actions. The interactor's actions are controlled by the interactor. Sensory information is passed from the physical world to the interactor through an interface controlled by a theory of presentation. In the project framework, the drama manager influences the characters' minds, the physical world, and the presentation theory. [23]

3.4 Emotion, Stress and Coping in Families with Adolescents: Assessing Personality Factors and Situational Aspects in an Experimental Computer Game (Geneva Emotion Research Group)

The project studies behavioral coping strategies developed by adolescents to face different types of stressful situations, with a specific focus on coping functionality, and complements coping research using questionnaires by controlled studies in the laboratory. Combining intra- and the inter-individual approaches to coping such that both situational and personality variables can be measured. Therefore, coping is studied in an intra-individual setting (one person confronted to different types of situations, at different moments in time), nested within an inter-individual setting (several individuals are compared with regard to their individual coping across different situations). [24]

3.5 The Cognition and Affect Project (University of Birmingham)

The main goal of this project is to understand the types of architectures that are capable of accounting for the whole range of human (and non-human) mental states and processes, including not only intelligent capabilities, such as the ability to learn to find your way in an unfamiliar town and the ability to think about infinite sets, but also moods, emotions, desires, and the like. For instance, they have investigated whether the ability to have emotional states is an accident of animal evolution or an inevitable consequence of design requirements and constraints, for instance in resource-limited intelligent robots. [25]

3.6 BlueEyes (IBM)

The project aims at creating computational devices with the sort of perceptual abilities that people take for granted. BlueEyes uses sensing technology to identify a user's actions and to extract key information. This information is then analyzed to determine the user's physical, emotional, or informational state, which in turn can be used to help make the user more productive by performing expected actions or by providing expected information. For example, a BlueEyes-enabled television could become active when the user makes eye contact, at which point the user could then tell the television to "turn on". [26]

3.7 People and Robot (CMU)

The project is directed at three little-understood aspects of service robots in society: the design and behavior of service robots; the ways that humans and robots interact; how service robots function as members of a work team. The initial domain for this work is elder communities and hospitals, where service robots can do useful but taxing tasks. The research aims at the design of appropriate appearance and interactions of service robots in these contexts. [27]

3.8 Affect Sensitive Human-Robot Collaboration (Vanderbilt University)

The projects involves developing a novel affect-sensitive architecture for human-robot cooperation, where the robot is expected to recognize human psychological states (for instance-stress, panic, fear, engagement in task at hand). This technique involves real-time monitoring of physiological signals of a human subject using wearable sensors. These may include his/her heart rate variability, brainwaves, skin conductance, respiration, muscle tension, blood pressure and temperature. The signals are analyzed in n real-time to infer the emotional states of the human interacting with the robot. The robot controller considers the psychological state in its feedback loop to decide on a course of action. The work exploits recent advances in control theory, signal processing, pattern recognition, and experimental psychology. [28]

3.9 Expressive Visual Speech Synthesis (NLPR, Institute of Automation, Chinese Academy of Sciences)

The project aims to enhance multimodal interfaces by adapting them to users' intentions and behaviours. For this, high-level characteristics of voices and faces are

defined that describe a person's expressiveness, as induced by the communicative intention, the current speaker state, the environmental condition, the relationship with his/her interlocutor(s), as well as by the person's identity, as circumscribed by gender, personality characteristics, age, language, and cultural membership. The project aims to make scientific breakthroughs by generating a multiplicity of voices and faces by extracting the features from speech, facial images and videos. The project will render man-machine communication more effective and natural by adding personality and expressiveness to the pure linguistic content that is currently generated by synthetic voices and talking heads, and by reacting to the intention and behaviour of the user.

4 Research Challenges

On the basis of perception, analysis, and modeling of affective features information such as speech and body language. The interrelation among research contents is illustrated in the following sketch:

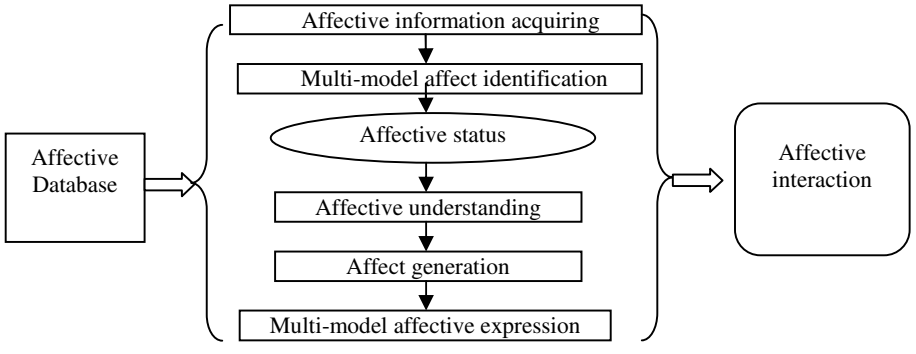

Fig. 1. Research framework

With the analysis above, some challenging research topics might be collected here.

4.1 Affective Understanding and Adaptation

Existing models of emotion use highly stylized stereotypes of personality types and emotional responsiveness, which do not correspond to real behavior in real people. There are lots of arguments in how to define the emotions. Someone might think it is not possible to model affect, and no way for affective understanding. This question has been discussed by Picard in her paper [63]. We know, "with any complex modelling problem, there are ways to handle complexity and contribute to progress in understanding. If we want to model affect at the neuropeptide signalling level, then we are limited indeed, because we know only a small part about how such molecules communicate among the organs in our body and realtime sensing of local molecular activity is not easily accomplished. "[63]

With the affect model, the ultimate purpose of affective computing is to assist the computer properly react after it understands the user's affect and meaning, and then be accustomed to the changes of the user's affect. Currently, there are some work use

the man-aided method to evaluate the user's feeling. It is still an important issue on how to analyze the dynamic characteristics of the user's affect and how to make the computer react properly according to the identification result of affective information. Affect is closed associated with personalities, environment, and cultural background, precise affect understanding model can only be realized by combining all these information. Psychological research results indicate that affect could be extended from the past affect states. Moreover, the lack of dynamic affect information mechanism is another important factor restricting current affect models. Therefore, how to define and integrate these information, how to describe/inegrate the dynamic affect information and how to improve the adaptation algorithm to natural scenarios should be the emphasis in the future research. It helps to build a personalized affective interaction system, by specifying the personal information and environment in real application.

4.2 Multi-model Based Affective Information Processing

As analysis in 2.4, the lack of the coordination mechanism of affective parameters under multi-model condition quite limits the affective understanding and the affect prompts. The amalgamation of different channels is not just the combination of them, but to find the mutual relations among all channel information. The mutual relation could make better integration of the different channels during interaction phases for both recognition/understanding and information generation. Figure 2 shows common affective status identification flow.

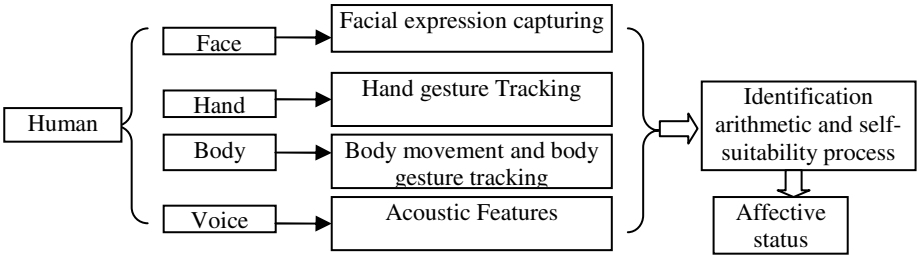

Fig. 2. Multi-model based affective recognition

4.3 Affective Feature Capturing in Real Environments

Most of current affective feature capturing is still limited in labs or studios, which are less complicated and have smaller background noises. The currently available information can only be used in information retrieval and common feature identification, which is too rough to make affective computing for complicated affect changes. Apart from developing high-quality affective interaction, we should emphasize on establishing automatic affective information capture from real environment and getting more reliable and detailed features, especially, for particular facial features' tracking and description, robust hand/body gesture tracking and modeling, more physiological acoustic parameters capture and modeling.

4.4 Affective Interaction in Multi-agent System

The study of agent based systems evolved from the field of Distributed Artificial Intelligence in the early to mid 1980's, with the development of intelligent Multi-Agent systems being given new impetus by the emergence of the World Wide Web and the Internet. Solving the problems associated with, and taking advantage of the opportunities offered by, this inherently distributed and unstructured environment are seen as major application areas for intelligent and Multi-Agent systems. Traditional affective interaction is just based on the single human computer interaction procedures. It is really a challenge work on how to make affective interaction in multi-agent system. In contrast to classical applications in artificial intelligence, the central ideas underlying multi-agent based affective interaction are that:

- the affect of one agent could be influenced by the other agents.
- the system exhibits goal directed behavior
- one agent can interact with and negotiate with other agents (possibly human) in order to achieve their goals
- the whole system can apply intelligence in the way they react to a dynamic and unpredictable environment.

Apart from the implementation of practical and useful systems another main goal in the study of Multi-Agent based affective interaction systems are to understand interaction among intelligent entities whether they are computational, human or both.

4.5 Affective Database

The shortage of affective database is one of the reasons why current study of affective computing is confined. Establishing a database storing a numerous number of affective data, especially multi-model affective data, is necessary to affective computing, and is also prerequisite for deeply studying affect mechanism. Some existing corpus consists of expressive speech [64], facial expression in videos [34][36][37][38], statistic 2D or 3D facial images [31][32], and motion capture data [30][43], etc. Nearly all of them are used for specific research, such as, emotion recognition, facial animation, etc. Due to lack of more detailed cognition experiments, the current database is hard to be used for the research of affective understanding. Further research should be involved in design, collection, marking, search, tool making, and other relative works related to multi-model affective data.

5 Conclusion

Though the concept of affective computing has come out not for a long time, it attracts high and extensive attention from academy and enterprise fields. The study and application of relevant fields are booming. To sum up, the existing researches are mainly limited in the detailed and scattered fields like voice and body language. Because of the lack of large affect data resources, no effective mechanism for multi-feature affective computing and relevant learning and controlling algorithms and the insufficiency of adaptation to natural scenarios, computers can not accurately judge and generate a human-like affect status and have real effective affect interaction. As a whole, various theoretical problems concerning affective computing are not well

solved. Even for that, there are still some applications, for instances, adding the function of automatic perception to people's mood in information household appliances and intelligent instruments to provide better services to people; making use of the function of affect concept analysis in computer retrieval system to improve the accuracy and efficiency of information retrieval; adding affect factors in the remote education platform may intensify the education effects; utilizing multi-model affect interaction technology in virtual reality application may build intelligent space and virtual scenario closer to real life etc. In addition, affect calculation may also be applied in the related industries like digital entertainment, robots and intelligent toys to realize more personalitive style and build more vivid scenario.

In recent years, the ubiquitous computing and wearable computing, which are closely related to affective computing, have achieved the pervasive attention of scientists. Ubiquitous computing and wearable computing are the necessary products of the combination of mobile computing technology and computer individualization. Concerning design, the computing technology becomes a part of our daily life and is closely tied with computer users. All these bring great conveniences to the real time capture of affect information as well as provide a perfect platform for affective computing. By means of the organic integration with affective computing, a colorful world of computing technology will be built.

References

[1] R.W. Picard, Affective computing. MIT Press, 1997.
[2] W. James, What is emotion? , Mind 9, 188-205, 1884.
[3] A. R. Damasio, Descartes, Error : Emotion, Reason, and the Human Brain, New York, NY: Gosset/Putnam Press, 1994.
[4] P. Ekman, Basic Emotions. Handbook of Cognition and Emotion. New York: John Wiley, 1999.
[5] H. Schlossberg, Three dimensions of emotion, Psychological review, 61, 81-88, 1954.
[6] C.E. Osgood,G.J. Suci,P.H. Tannenbaum, (eds.), The measurements of meaning, University of Illinois Press, 1957.
[7] Sylvie J.L. Mozziconacci and Dik J. Hermes, Expression of emotion and attitude through temporal speech variations, ICSLP2000, Beijing, 2000.
[8] J.E. Cahn, The generation of affect in synthesized speech, Journal of the American Voice I/O Society, vol. 8, July 1990.
[9] N.L. Etcoff, J.J. Magee, Categorical perception of facial expressions, Cognition, 44, 227-240, 1992.
[10] A.Camurri, G.De Poli, M.Leman, G.Volpe, A Multi-layered Conceptual Framework for Expressive Gesture Applications, Proc. Intl MOSART Workshop, Barcelona, Nov. 2001
[11] Jianhua Tao, Emotion Control of Chinese Speech Synthesis in Natural Environment. Eurospeech2003, Geneva, Sep.2003
[12] R. Cowie., Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18(1):32-80, 2001.
[13] A.Azarbayejani, etc., Real-Time 3-D Tracking of the Human Body, IMAGE'COM 96, Bordeaux, France, May 1996
[14] J. F. O'Brien, B. Bodenheimer, G. Brostow, and J. Hodgins, ``Automatic Joint Parameter Estimation from Magnetic Motion Capture Data", Proceedings of Graphics Interface 2000, Montreal, Canada, pp. 53-60, May 2000.

[15] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang , Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997
[16] D. M. Gavrila, The Visual Analysis of Human Movement: A Survey, Computer Vision and Image Understanding, Vol. 73, No.1. January, 1999: 82-98
[17] J. K. Aggarwal, Q. Cai, Human Motion Analysis: A Review, Computer Vision and Image Understanding, Vol. 73, No. 3, 1999
[18] Tsuyoshi Moriyama, Shinji Ozawa, Emotion Recognition and Synthesis System on Speech, IEEE International Conference on Multimedia Computing and Systems, 1999 Florence, Italy
[19] R. Antonio, and Hanna Damasio. Brain and Language. Scientific American September, 1992. 89-95
[20] A. J. Calder, A Principal Component Analysis of Facial Expression. Vision Research, 2001, Vol 41.
[21] http://emotion-research.net/
[22] http://affect.media.mit.edu/
[23] http://www.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/oz.html
[24] http://www.unige.ch/fapse/emotion/
[25] http://www.cs.bham.ac.uk/%7Eaxs/cogaff.html
[26] http://www.almaden.ibm.com/cs/BlueEyes/index.html
[27] http://www.peopleandrobots.org/
[28] http://robotics.vuse.vanderbilt.edu/affect.htm
[29] Dominic W. Massaro , Jonas Beskow, Michael M. Cohen, Christopher L. Fry, and Tony Rodriguez. Picture My Voice: Audio to Visual Speech Synthesis using Artificial Neural Networks. Proceedings of AVSP'99, pp.133-138. Santa Cruz, CA., August, 1999.
[30] Yamamoto E., Nakamura, S., & Shikano, K. Lip movement synthesis from speech based on Hidden Markov Models. Speech Communication, 26, (1998).105-115
[31] R. Gutierrez-Osuna, P. K. Kakumanu, A. Esposito, O. N. Garcia,A. Bojorquez, J. L. Castillo, and I. Rudomin. Speech-Driven Facial Animation With Realistic Dynamics. IEEE Trans. on Multimedia, Vol. 7, No. 1, Feb, 2005
[32] Pengyu Hong, Zhen Wen, and Thomas S. Huang, Real-time speech-driven face animation with expressions using neural networks. IEEE Trans on Neural Networks, Vol. 13, No. 4, July, 2002.
[33] A. Murat Tekalp, , JoK rn Ostermann, Face and 2-D mesh animation in MPEG-4, Signal Processing: *Image Communication* 15 (2000) 387-421.
[34] Bregler, C., Covell, M., Slaney, M., Video Rewrite: Driving Visual Speech with Audio, ACM SIGGRAPH, 1997.
[35] Hunt, A., Black, A., Unit selection in a concatenative speech synthesis system using a large speech database, ICASSP, vol. 1, pp. 373-376, 1996.
[36] Cosatto E, Potamianos G, Graf H P. Audio-visual unit selection for the synthesis of photo-realistic talking-heads. In: IEEE International Conference on Multimedia and Expo, ICME 2000. 2: 619~622
[37] T. Ezzat, T. Poggio, MikeTalk: A Talking Facial Display Based on Morphing Visemes, in Proc. Computer Animation Conference, Philadelphia, USA, 1998.
[38] Ashish Verma, L. Venkata Subramaniam, Nitendra Rajput, Chalapathy Neti, Tanveer A. Faruquie. Animating Expressive Faces Across Languages. IEEE Trans on Multimedia, Vol. 6, No. 6, Dec, 2004.
[39] P. Ekman and W. V. Friesen, Facial Action Coding System. Palo Alto, Calif: Con

[40] Nick Campbell, "Perception of Affect in Speech - towards an Automatic Processing of Paralinguistic Information in Spoken Conversation", ICSLP2004, Jeju, Oct, 2004.
[41] Andrew Ortony, Gerald L. Clore, Allan Collins, "The Cognitive Structure of Emotions", book
[42] C. Gobl and A. N´ı Chasaide, "The role of voice quality in communicating emotion, mood and attitude," Speech Communication, vol. 40, pp. 189–212, 2003.
[43] Scherer K.R., "Vocal affect expression: A review and a model for future research," Psychological Bulletin, vol. 99, pp. 143–165, 1986.
[44] E. Eide, A. Aaron, R. Bakis, W. Hamza, M. Picheny, and J. Pitrelli, A corpus-based approach to <ahem/> expressive speech synthesis, IEEE speech synthesis workshop, 2002, Santa Monica
[45] Dellaert, F., Polzin, t., and Waibel, A., Recognizing Emotion in Speech", In Proc. Of ICSLP 1996, Philadelphia, PA, pp. 1970-1973, 1996.
[46] Petrushin, V. A., "Emotion Recognition in Speech Signal: Experimental Study, Development and Application." ICSLP 2000, Beijing.
[47] Lee, C.M.; Narayanan, S.; Pieraccini, R., Recognition of Negative Emotion in the Human Speech Signals, Workshop on Auto. Speech Recognition and Understanding, Dec 2001.
[48] Tato, R., Santos, R., Kompe, R., Pardo, J.M., Emotional Space Improves Emotion Recognition, in Proc. Of ICSLP-2002, Denver, Colorado, September 2002.
[49] Yu, F., Chang, E., Xu, Y.Q., and Shum, H.Y., Emotion Detection From Speech To Enrich Multimedia Content, in the second IEEE Pacific-Rim Conference on Multimedia, October 24-26, 2001, Beijing, China.
[50] Nick Campbell, Synthesis Units for Conversational Speech - Using Phrasal Segments, http://feast.atr.jp/nick/refs.html
[51] M. Schröder & S. Breuer. XML Representation Languages as a Way of Interconnecting TTS Modules. Proc. ICSLP'04 Jeju, Korea.
[52] Ze-Jing Chuang and Chung-Hsien Wu "Emotion Recognition from Textual Input using an Emotional Semantic Network," In Proceedings of International Conference on Spoken Language Processing, ICSLP 2002, Denver, 2002.
[53] H. Kobayashi and F. Hara, "Recognition of Six Basic Facial Expressions and Their Strength by Neural Network," *Proc. Int'l Workshop Robot and Human Comm.*, pp. 381-386, 1992.
[54] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi , Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets. *Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition,* April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.
[55] R. Brunelli and D. Falavigna, ''Person Identification Using Multiple Cues'', *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 12, No. 10, pp. 955-966, Oct 1995.
[56] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer, ''Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics'', Proc. *International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA)*, pp. 291-300, Crans-Montana, Switzerland, March 1997.
[57] A. Kumar, D. C. Wong, H. C. Shen, and A. K. Jain, ''Personal Verification using Palmprint and Hand Geometry Biometric'', *4th International Conference on Audio- and Video-based Biometric Person Authentication*, Guildford, UK, June 9-11, 2003.
[58] R. W. Frischholz and U. Dieckmann, ''Bioid: A Multimodal Biometric Identification System'', *IEEE Computer*, Vol. 33, No. 2, pp. 64-68, 2000.

[59] A. K. Jain and A. Ross, ‘‘Learning User-specific Parameters in a Multibiometric System’’, Proc. International Conference on Image Processing (ICIP), Rochester, New York, September 22-25, 2002.
[60] T. K. Ho, J. J. Hull, and S. N. Srihari, ‘‘Decision Combination in Multiple Classifier Systems’’, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 66-75, January 1994.
[61] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, ‘‘On Combining Classifiers’’, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp. 226-239, Mar 1998.
[62] U. Dieckmann, P. Plankensteiner, and T. Wagner, ‘‘Sesam: A Biometric Person Identification System Using Sensor Fusion’’, *Pattern Recognition Letters*, Vol. 18, No. 9, pp.827-833, 1997.
[63] Picard, RW, Affective Computing: Challenges, Int. Journal of Human-Computer Studies, Vol. 59, Issues 1-2, July 2003, pp. 55-64.
[64] Nick Campbell, Databases of Expressive Speech, COCOSDA 2003, Singapore
[65] http://www.ubiq.com/hypertext/weiser/UbiHome.html