



An Intelligent Robot Based on Emotion Decision Model

Liu Yaofeng* **Wang Zhiliang** **Wang Wei** **Jiang Xiao**

School of Information, Beijing University of Science and Technology, Beijing 100083, China

*Corresponding author's E-mail: yaofeng-liu@163.com

Abstract: This study discusses the development of an intelligent robot, its emotion decision model and its behavior expression model. According to bionics and facial action code system, six basic facial expressions of the robot are realized using multi-motors cooperative control. On the grounds of linear dynamic system and personality matrixes, emotional model of the robot consists of five elements: reaction dynamic, internal dynamic, emotional dynamic, behavior dynamic and personality. Therefore, the robot uses machine vision, voice recognition and infrared sensors which could make it perceives surroundings. Moreover, it also could detect face, analysis emotion in voice and decide actions intelligently. With emotional interactive model, the robot could interact with human being friendly and harmoniously.

Keywords: Intelligent robot; Human-robot interaction; Emotion interaction model

1. Introduction

The combination of smart machines and emotion has been a research topic for decades. The realization of emotion is the key point for smart machines to interact with people. Robot is able to read people's emotion to communicate with them naturally. And robot perceives user's face, body, language and eyes through its vision. In addition, understand the emotion of user is also important. So robot possessed with human emotion has many interactive methods to express thief emotion, such as face, spinning eyes, voice and body language.

At present, the research of emotional robots is mainly in universities and research institutions of the Unites States, Japan and the European Union. Cynthia Breazeal from Artificial Intelligence Laboratory at the Massachusetts Institute of Technology has developed an expressive anthropomorphic robot called Kismet [1, 2]. Hanson Robotics of America has introduced variety of robots with facial expressions [3, 4]. Waseda University, Japan, has developed a series of robots named WE-R (four in total) since 1996[5, 6]. A project focusing on humanoid robot heads has been launched in University of Kaiserslautern, German

[7].In the following, an emotional robot with facial expressions which could communicate with persons naturally through facial expressions, computer vision and speech is introduced. The emotional robot detects the external environment through infrared sensors, microphones and USB cameras. Information from infrared sensors is processed by PIC, and passed to PC through serial port. Information from microphones and USB cameras is passed to PC directly. Infrared sensors are used to detect external obstacles; microphones are for speech input and USB cameras for human face detection. Emotional features are firstly extracted from external information by PC, and then voice emotion and facial expressions of detected face are calculated. The external information is finally synthesized into the emotion that robot will express. Voice output module, the body language of the mechanical head and the facial expressions constitute the robot's way of emotional expression.

2. Development of the humanoid robot head

The production of the humanoid robot head can be divided into four steps: first, design of

three-dimensional mechanical structure, CAXA entity 2006 used in this research; second, processing and assembling of mechanical structure, hard Aluminum(LY12) used for most parts to ensure light weight and high strength; Third, installation of fiberglass shell, eyeballs and eyelids; fourth, production and installation of silicone skin, facial expression units[8, 9], hair, eyebrows and lash, color of eyeballs and face.

2.1 Design of mechanical structure of head

The whole bracket of head was processed with duralumin, which could support the head. Brow and mouth steering engines are installed on it. Moreover, there are eyes backplane and controlling circuit of steering engines. Manufacture of expressional robot head has four steps.

- 1) Three-dimensional design of mechanical structure: using CAXA entity 2006;
- 2) Processing and assembling of mechanical structure using duralumin (LY12) material;
- 3) Fix of fiberglass crust, eyeball and eyelid;
- 4) Installation of silicone skin, facial expressional action, hair, eyelash and eyebrow. Colored eye and make up face.

2.2 Electric motor selection

Space in robot's head is very small. But 13 degrees of freedom motion are needed. Because steering engine is of features that compact size, easy installation, large torque output, good stability, simple controlling, convenient linking with digital system, it is used as power-drive of the robot's head. And its model is GWS815. Compared with steering engines of the same type, GWS has bigger torque. In addition, it is easy to control and stable. Steering engine (HG14-M) which has bigger torque and volume is adopted to drive the neck of robot. Robot has actions of vertical head nod and lateral headshake which are driven by two steering engines separately. And the other 11 motors help cheek of robot act.

2.3 System structure

This article also studies on electric control system of robot as hardware foundation. Whole system adopts a structure of two layers. Top layer (PC) processes images, recognizes and synthesizes voices which have very high computational

complexity. Bottom layer (PIC16F877) is SCM. Two layers are connected with RS-232 serial port or wireless module to communicate. Bottom layer receives and processes information from sensors. And it also drives motors and controls the robot.

System chart of robot's head is shown as Figure 1. The robot apperceives surroundings through machine sight, voice recognition and IR sensors. Machine sight refers to facial detecting function. Voice recognizing module transmits voice to the host computer with microphone. And it also extracts emotion from voice. IR sensors could find out whether there are obstacles around robot or human being near by. Then transmit the signal detecting to the slave computer system. After processing, it is got by the host computer through RS232. Considering image, voice and slave computer signal, information of surroundings and human being interacting with robot could be got. Emotion output module transmits instruction to slave computer system and voice synthesis module. In addition, slave computer controls motors to produce body languages and facial expression through PWM. Synthesis voice is shown as language of robot using amplifier. Therefore, voice, body language and facial expression make up robot emotion together. Every function module could be adjusted by using debugging interface in the host computer.

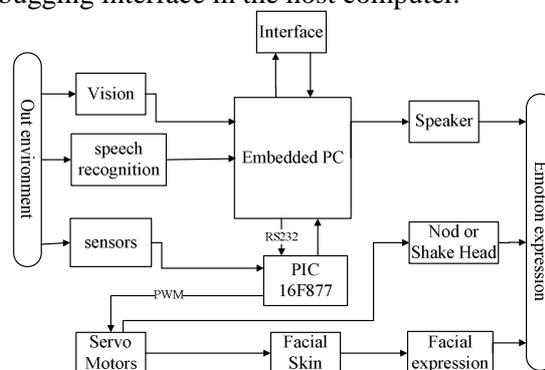


Figure.1 Robot's architecture.

2.4 Facial expression debugging

As shown in Figure 2, click button Browse, read head movement data of robot. CH1-CH13 are corresponding to 13 steering engines. In the situation of robot with power, corresponding steering engine of robot could turn certain angle by dragging any slider in CH1-CH13. According to the active needs of robot, every steering engine is set to turn certain angle. So the robot could finish some actions. This software also has fine turning function. Select one group of values, double click, and they are shown on sliders. By dragging slider, every

value corresponding to CHX could be changed. After that, the new value is updated and saved by clicking the button Modify on the top-right area of the software. And on the top of interface, there is a synchronous option shown as ON or OFF. On the one hand, if OFF is selected, values which are adjusted completely are sent to robot by clicking the button Transmit laid alongside the button Insert. These values correspond to the 13 sliders. So whether values are adjusted completely is certified. On the other hand, robot acts synchronously when dragging sliders. That is to say fine turning values are shown in real time.

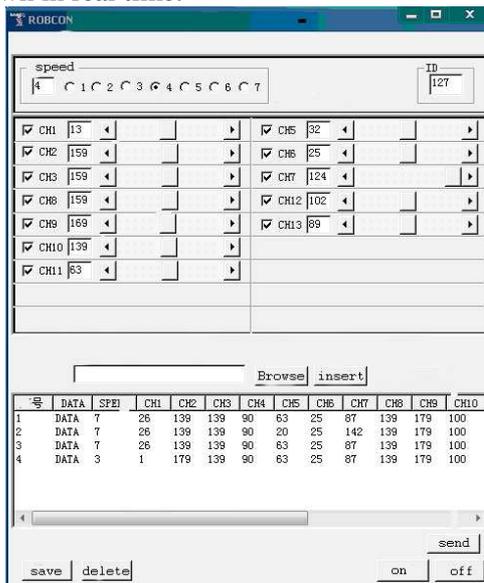


Figure.2 Debugging interface.

3. Facial detection

On condition that background was simple and static, meanwhile angle of head was no more than 20 degrees to left or right; the detection rate could be higher than 98% through experiments. In addition, the error rate was almost 0%.

About 56 milliseconds were spent on detecting a single image which is of 320*240 pixels in common video. The FPS (Frames per Second) of normal video is about 24-30, that is to say a single frame in video is 33-42ms. In order to have enough time in later processing for real-time, such as facial recognition and expression recognition, the time cost by detecting algorithm should be less than 20ms.

It is suitable to contract the detection region and extend the range of minimum facial scale to reduce the time consumed by Haar feature extracting. Two continuous frames are related to each other in video. So this feature of timing sequence could be used to reach the goal above. An improved facial detecting algorithm used in video sequence is presented soon

after. It is based on Haar feature. Supposing that there exist only one face in video which will be detected, the detecting steps are as follows:

Step A: If it is the first frame or there is no face in previous frame, go to Step C. Or else enlarge face detecting region by 1.2 times than the one in previous frame. At the same time, minimum facial scale reduces by 0.8 times. For instance, while the minimum facial scale in previous frame is 100*100, the one in current frame is (100*0.8)*(100*0.8). Operation on current frame is finished if face is found.

Step B: Calculate difference of gray scale image between current and previous frame. Then image binarization with adaptive threshold is carried out. There exist some improvements in this step. Selecting two images in continuous frame of video. After image binarization, changing region could be got through vertical and horizontal projection. According to this region, new one is deduced by following the rules in Step A. Then detecting face in this new region. To 320*240 pixels image, time cost by this step is about 4ms through experiments.

Step C: Detect face in the whole image. And minimum facial scale is set to 20*20. If face is found, operating on current frame is finished; otherwise there is no face in current frame.

The steps above are in accordance with three face states in video. The first state is that face detected moves a little in camera area. That is to say priori knowledge diminishes the area detected. So with Step A, face could be found. The second state is that face detected moves commonly. Method proposed in Step B could diminish the area detected. Then Haar feature extracting is help to detect face in real time. The third state is face moves a lot. Step C is effective. Compared with detecting face in the whole frame, steps above seem a little complex. But face detecting speed improves obviously for video. If face exists in video, the position and facial scale between current and previous frame are of correlativity. Based on priori knowledge of some previous frames, detecting area and facial scale are restricted to speed up algorithm.

Facial detecting program runs with ordinary camera and notebook (CPU: Pentium 1.73G; Memory: 512MB). Its environment needs VC++ 6.0 and Open CV. For frame which has 320*240 pixels, average time cost by processing is about 10ms through experiments. It shows Haar feature extracting is of high detection rate and low error rate. Meanwhile, optimizing method above is practical. Effects of facial detecting are shown as Figure 3. Area which is same as complexion is not detected. Figure 4 shows that the algorithm is robust in lighten

environment, which guarantees correct facial detecting. So with this optimized algorithm, subsequent facial recognition has foundation for real-time in video.

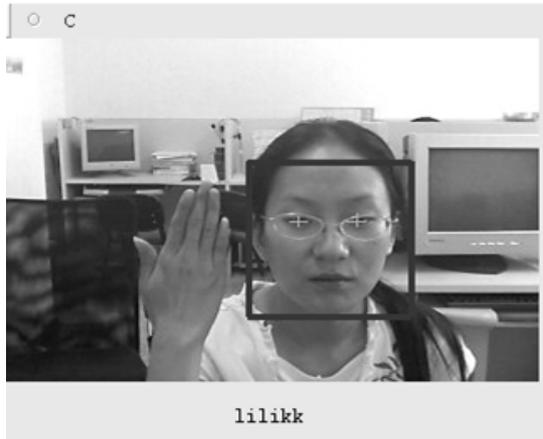


Figure.3 Facial detect image.



Figure.4 Facial detect image under different illumination.

4. Speech recognition and synthesis

Module of speech recognition in robot is developed with Pattek ASR SDK and VC++ 6.0. Pattek ASR SDK is easy to use. In addition, it has high recognition rate and no need to train. And with the design of its engine, it can be used by non-special persons meanwhile the users can define vocabularies and grammars. The flowchart of the module above is shown in Figure 5.

Module of speech synthesis in robot is developed with iFly TTS SDK and VC++ 6.0. It is of high quality in text-to-voice transmitting. Large-scale corpus synthesis technology and very large scale natural speech database are adopted. Therefore data statistics and machine learning could be done with these. Based on linguistics, phonetics and artificial intelligence knowledge, models of

fundamental frequency, period, energy and pause could be set up. To get the model, decision tree and Nero network in machine learning are utilized. Meanwhile, module is endowed with highly efficient integrated strategy. Synthesizing and playing is simultaneous to avoid long time lag in synthesizing long text. When playing voice, any break by the user could stop current synthesizing. This method diminishes meaningless resource consuming. At the same time, speaking speed, synthesizing style and sound volume could also be adjusted. Synthesizing voice is continuous, understandable and natural just like human speak. Flow chart of this module is as Figure 6.

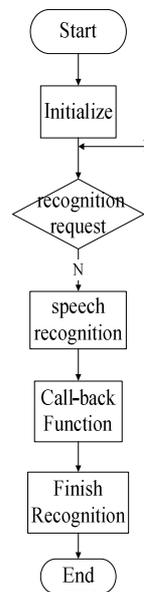


Figure.5 Speech recognition.

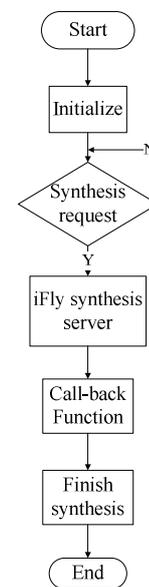


Figure.6 Speech synthesis.

4.1 Emotion analysis in robot conversation

According to the user's voice, robot could feel and find the corresponding emotion, then synthesize voice with it.

4.2 Emotion classification in robot conversation

In modern psychologists, emotion is grouped into four fundamental classes: happy, angry, sorrow and fear. Then they are divided into six ranks showed in Table 1.

Table 1. Emotion classification and ranks

ranks	happy	angry	fear	sorrow
1	satisfaction	angry	trepidation	etdown
2	cheer	chafe	afraid	sad

3	exaltation	outrage	panic	sorrow
4	happy	rage	fear	rueful
5	exultation	fury	affright	glum
6	ecstasy	wingding	terror	grieved

In this paper, robot emotion is classified into five types: happy, angry, sorrow, fear and normal. Based on this, we distinguish patterns from emotion.

4.3 Emotion analysis of input voice

Treatments of different emotion are adopted related to five types above. Happy, angry, sorrow and normal could be treated as the same. If dialogs input by the user are commendatory, robot emotion is defined as happy. If they are average derogatory sense, robot emotion is sorrow. If they are strong derogatory sense, emotion is angry. And if they are objective, emotion is normal. According to feature of common kids, something special which children are afraid of is adopted to result in fear emotion. Special things include ghost, wolf, knife and so on.

4.4 Emotion conversion in robot conversation

Emotion expressed by robot includes happy, angry, sorrow, fear and normal. Quantify the emotion and show in Table 2 and Table 3.

Table 2. Robot emotion quantification 1

Emotion	trepidation	afraid	panic	fear
Quantification	-6	-5	-4	-3
Emotion	affright	terror	normal	etdown
Quantification	-2	-1	0	1
Emotion	sad	sorrow	rueful	dump
Quantification	2	3	4	5
Emotion	grieved	angry	chafe	outrage
Quantification	6	7	8	9
Emotion	rage	fury	wingding	
Quantification	10	11	12	

Quantification profits robot to convert emotion easily. Such quantification guarantees that there is no overlap when happy and fear with the same quantified value inquire. In this paper, normal emotion is standard. If robot has happy or fear emotion at the beginning, it converts into normal emotion first. Then it carries out next convention. In

the other side, if current emotion is fear, it turns normal and determines next emotion.

Emotion converting adopts medium value. That is if current emotion of input voice is m_input and the previous emotion of robot is $m_prefeeling$, the output emotion is m_output . Generally speaking, a formula is deduced as:

$$m_output = \left\lfloor \frac{m_input + m_prefeeling}{2} \right\rfloor$$

Table 3. Robot emotion quantification 2

Emotion	normal	satisfaction	cheer	exaltation
Quantification	0	-6	-5	-4
Emotion	happy	exultation	ecstasy	etdown
Quantification	-3	-2	-1	1
Emotion	sad	sorrow	rueful	dump
Quantification	2	3	4	5
Emotion	grieved	angry	chafe	outrage
Quantification	6	7	8	9
Emotion	rage	fury	wingding	
Quantification	10	11	12	

And influence from converting to normal emotion could be thought in practice.

With the model above, next emotion could be deduced from current input emotion and previous emotion, which guarantees smooth conversion. Emotion converting is shown in Figure 7.

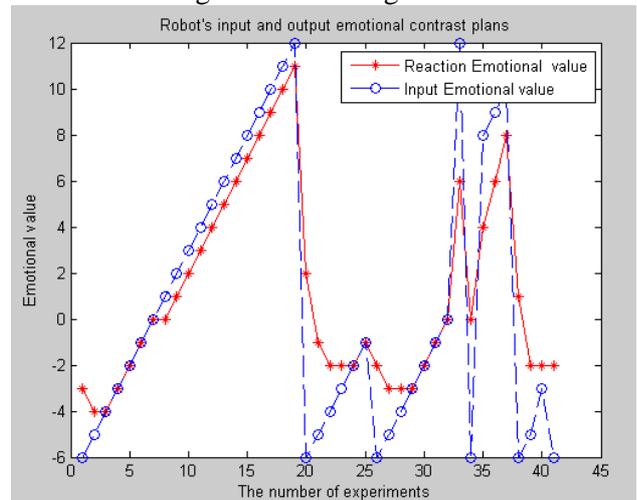


Figure.7 Comparison between input and output using emotion converting model
As shown in Figure 7, output emotion of robot is

indicated with red-stars line, while input emotion is blue-circles line. At the very beginning, robot has default emotion: normal. That is to say default value is 0. Because of adopting medium value, output emotion is smoother than input one. Relatively, transitive process is gentle. But without the model above, output emotion equal to input one. In short, while value changes smoothly, emotion converts smoothly too. Gradual conversion is in accordance with feature of human's emotion.

Firstly, determine which emotion robot should express in conversation with model above. Then express it in synthesis voice.

5. Robot behaviors expression

5.1 Basic facial expression

According to various facial expressions in debugging process, expression library was set up. There are surprise, fear, dislike, angry, happy and sorrow emotion in it. They are shown in Figure 8.



Figure.8 Facial expression.

Using various basic facial expression, robot could convey different emotion. For example, mournful action demonstrates robot is glum or unwilling. If it wants to express excited emotion, happy or respective facial expression would be used. In the same way, robot could show surprise, fear, angry and happy. Along with emotional synthesis voice, robot emotion expression is more vivid.

5.2 Model of emotion decision

Personality is one of the important characteristics in human-computer interaction. This point makes robot interact with human being in friendly way instead of making them feel uncomfortable. Model of emotion decision has been adopted to help our robot have more personality and characteristic. With this model, robot decides which action it should take by itself. In addition, the most

important feature of robot is changeable personality. When it changes, robot would act differently in the same situation. Here shows the structure of model of emotion decision in Figure 9. It consists of five elements: reaction kinetics, internal dynamic, emotional dynamic, behavior dynamic and personality. Emotion model is on the basis of linear dynamic system and various personality matrixes. While robot personality is different, its decision is also different to the same external stimulus and internal power.

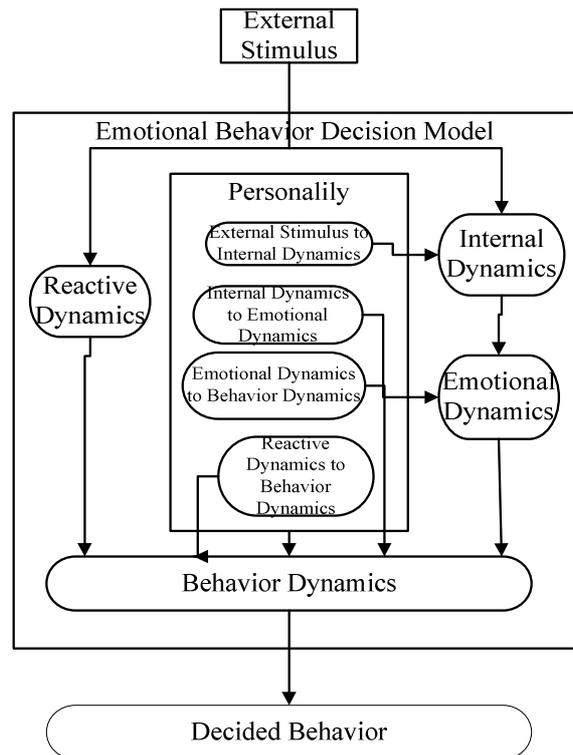


Fig.9 Emotional Behavior Decision Model

Model effects on reaction kinetics and internal dynamic when external stimulus exists. Reaction kinetics makes robot respond to external stimulus. But it is not influential in emotional dynamic. For example, robot may be sad or excited under stimulus. Moreover behavior could also be seen when robot recognizes faces. That is to say robot shows excited at familiar faces, while scared or confused at suspicious ones. External stimulus includes facial recognition, gesture recognition, and voice recognition and so on. In the emotional behavior decision model, internal dynamic simulates consciousness effect. It determines if external stimulus should influence on robot emotion state. With it, robot could recognize whether stimulus is a joke or something caused anger. Emotional dynamic rests with internal dynamic and personality. When there is stimulus, factors in system are all

recalculated. Furthermore, the renewal process is related to previous data.

Reaction kinetics, internal dynamic, emotional dynamic and behavior dynamic are dependent on personality matrixes. So while personality is changing, robot would act differently in the same situation. Final behavior is determined by behavior dynamic. Every behavior has a value. It changes with external stimulus and internal dynamic. Then select usable behavior with highest value and behavior with threshold value. If the value of final selected behavior is less than the value of maintaining original behavior, stay still.

Personality of robot is easy to change with modified personality matrix. So robot structure is changeable for various applications.

6. Experiment result and analysis

To evaluate effects of emotion expressed by robot, two experiments are designed. One is simplex facial expression; the other is fusion of output voice and facial expression. Distance between participator and robot is about three meters in experiments. Participator spends 3-5 seconds on observing every emotion expressed by robot. And total number of participator is 60. Every participator records his observation of robot emotional state. Statistics are shown in Table 4 and Table 5. Ordinate indicates emotion state expressed by robot. Abscissa is emotion thought by participator.

Table. 4 Evaluation results of simplex facial expression

	Ang er	Disg ust	Fe ar	Happ iness	Sad ness	Surp rise
Anger	20	1	8	10	1	20
Disgust	4	21	1	18	15	1
Fear	0	0	30	12	3	25
Happiness	0	0	3	40	2	15
Sadness	8	20	1	0	31	0
Surprise	0	0	11	20	0	29

Table. 5 Evaluation result of fusion of output voice and facial expression

	An ger	Dis gust	Fe ar	Happ iness	Sad ness	Surp rise
Anger	35	0	0	15	0	10
Disgust	0	40	0	3	17	0
Fear	0	0	35	10	5	10
Happiness	0	0	0	50	2	8
Sadness	2	15	1	0	43	0
Surprise	5	0	15	15	0	35

According to analysis of Table 4 and Table 5, comparisons among simplex facial expression, fusion of output voice and facial expression show robot emotion more truly.

7. Conclusion

Applications of intelligent robot are changing greatly along with artificial intelligence developing as time goes on. Robot is becoming a part of members of our society gradually. In the past few days, progress of robot integration was a little slower. But with the development of new technology, robot becomes more and more intelligent. Therefore, in our society, robot will be seen everywhere at all times.

In this article, a type of robot which is bright enough to adapt itself to the new condition is proposed. It could recognize person by detecting and recognizing face. In addition, it also could express its emotion and personality with decision model of emotional action. Furthermore, it could be applied to lots of conditions in practice. It can be used as family service robot, elders-taken-care robot, waiters in restaurant, nurse in hospital, exhibition and market instructor. Generally, the research focuses on interacting of human being and robot. And the interacting should be friendly and vivid. Moreover, results show that our robot could express many actions according to current situation. If interaction with person has changed, the robot can adapt itself to the changing. In the future, something should be done such as enhancing the capability of the robot and increasing objects recognizing, sensing, tracking and monitoring, which makes robot can be applied in multi applications.

Acknowledgments

This work was supported by The 863 National High-tech Development Plan of China No. 2007AA01Z160 and The Beijing Natural Science Foundation of key Projects No. KZ200810028016.

References

- [1] Cynthia Breazeal, *Designing Sociable Robots*, The MIT Press, 2002.
- [2] Cynthia Breazeal, *Toward Sociable Robots*, Robotics and Autonomous Systems, 2003.
- [3] Hanson, D. Rus, D., Canvin, S., Schmierer, G., Biologically Inspired Robotic Applications, *In Biologically Inspired Intelligent Robotics[C]*, SPIE Press, pp.394-399. 2003.
- [4] Hanson, D, Olney, A, Zielke, M, Pereira, A. Upending

- the Uncanny Valley[C], *AAAI conference proceedings*, pp.411-418. 2005.
- [5] Hiroyasu Miwa, Tomohiko Umetsu, Atsuo Takanishi, Human2 like robot head that has olfactory sensation and facial color expression. *Proceeding of the 2001 IEEE[C]*, ICRA, pp.459 - 464. 2001.
- [6] A. Takanishi, H. Miwa, and H. Takanobu, .Development of humanlike head robots for modeling human mind and emotional human-robot interaction,. *IARP International workshop on Humanoid and human Friendly Robotics*, pp. 104.109, December 2002.
- [7] Karsten.Berns,Jochen Hirth, Control of facial expressions of the humanoid robot head ROMAN ,*Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*,Beijing,China, pp.504-509. 2005.
- [8] P. Ekman and W. Friesen, Facial Action Coding System. *Consulting psychologist Press, Inc*, 1978.
- [9] P. Ekman, W. Friesen, and J. Hager, Facial Action Coding System. *A Human Face*, 2002.