Affective Interaction Framework for Humanoid Robots

한국과학기술원 양현승 · 서용호 · 정일웅

1. Introduction

To interact sociably with a human, a robot or a computer not only has to recognize and express emotions like a human, but also has to share knowledge with humans. It also has to speak and determine its autonomous behavior while considering its memory and the emotional status of a human. In this section, we **present an affective interaction system between human and robot.** This system **communicates with humans by sharing experiences and knowledge on users and topics.** Interaction between human and robot is made possible through the framework presented in this section. The framework enables a robot to catch emotional status of current user and to activate its appropriate memories. As a result, the robot naturally engages in dialogue with a human; it chooses appropriate conversation topics and behaves appropriately in response to its memory and human emotions. Moreover, the human partner perceives the robot to be more human like and friendly, thus enhancing the interaction between the robot and human.

2. Affective Interaction Framework

We designed the affective interaction framework to include the five subsystems shown in Figure 1.

The arrows in Figure 1 represent the information flow and influence between the subsystems. Well coordinated communication among these subsystems is required for our robot to successfully function in a dynamic world.

The main functions of each subsystem are summarized as follows. The perception system is

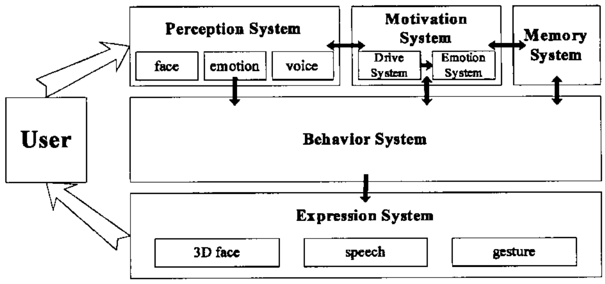


Figure 1 Affective Interaction Flow

a system that mainly extracts information about outside worlds. In our system, the perception system detects whether the human is there or not, who is the human, how is his emotional status, and what he says. The perception system is composed of the sub-systems such as face detection, face recognition, emotion recognition, and motion & color detection.

The motivation system is composed of drive and emotion system. Drives are motivators, and include endogenous drives and externally induced desires. The robot has three basic drives for accomplishing its tasks and objectives: the drive to interact with a human, the drive to ingratiate itself with a human and the drive to maintain its well-being. The emotion synthesis system makes the robot's artificial emotions. Currently, three emotions were modeled to give its synthetic analogues of anger, joy, and sorrow.

The memory system has two main functions: First, the memory system saves the most frequently occurring emotion in the latest interaction with a user, thus, influencing the robot's emotional status when meeting the person again. Secondly, the memory system maintains memories of previously mentioned topics, and periodically increases knowledge on it by searching the keyword in the Internet.

Next, the behavior system performs action selection i.e., it chooses the most relevant behavior to perform given the perception, motivation and memory input. Lastly, the expression system plays expressions composed of 3D facial the expressions, dialog, and gestures according to the result of the behavior system.

3. Perception System

The perception system comprises face detection, face recognition, speech recognition, emotion recognition, and visual attention detection that enables the robot to detect objects and color. Accordingly, through the perception system, the robot can learn basic knowledge such as the identity of the human user, the user's emotional state, and what the user is saying and doing. The overall process of the perception system is shown Figure 2.

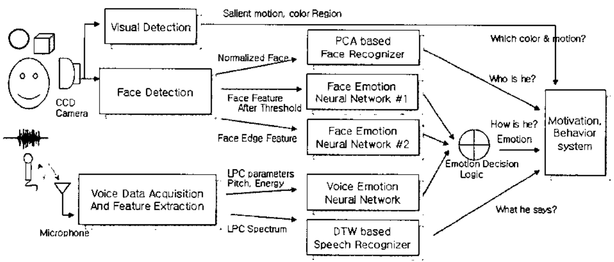


Figure 2 Architecture of the Perception System

4. Bimodal Emotion Recognition

We estimate emotion through facial expression and speech, and then integrate them to enable bimodal emotion recognition.

For emotion recognition through facial expression, we normalized the image captured. We then extracted the following two features, which are based on Ekman's facial expression features [1]: One is facial image of lips, brow and forehead, and the other is edge image of lips, brow and forehead.

For emotion recognition through speech, we adopted a recognition method similar to the one used in life-like the communication agents called MUSE and MIC[2]. Of the system's two features, one is a phonetic feature and the other is a prosodic feature. We trained each feature vector using a neural network.

For bimodal emotion recognition, we integrated two training results using decision logic. Final result vector of the decision logic R_{final} is as follows:

$$R_{final} = (R_{face} W_f + R_{speech} W_s) + R_{final-1} - \delta t$$

R_{face} and R_{speech} are the results vector of the emotion recognition through facial expression and speech, W_f and W_s are the weights of the two modalities, $R_{final-1}$ is the previous emotion result determined by decision logic, and δt is a decay term that eventually restores the emotional status to neutral.

Our emotion recognition system yielded approximately 80 percent for each of five testers. By resolving confusion, it achieved better performance than facial-only and speech-only systems.

5. Motivation System

The motivation system sets up the robot's nature by defining its "needs" and influencing how and when it acts to satisfy them. The nature of the robot is to affectively communicate with humans and ultimately to ingratiate itself with them. The motivation system consists of two related subsystems, one that implements drives and a second that implements emotions. Each subsystem serves as a regulatory function for the robot to maintain its "well-being".

In the drive system, three basic drives were defined for the system objective, the affective interaction with a human: the drive to interact with a human; the drive to ingratiate itself with a human and drive to maintain its well being. In the current implementation, there are three drives

- To interact with humans: This drive motivates the robot to find and approach to a human, and greet him. Accordingly, if the robot cannot find and greet a human through face detection and recognition of the perception system, its activation intensity increases.

- To ingratiate itself with humans: This drive prompts the robot to make the human feel better. When the robot interacts with a human, it tries to ingratiate itself while considering the person's emotional state. When the person feels sorrow, the robot attempts to console the person; when the person feels surprise or anger, the robot attempts to pacify the person. If the intensity of the recognized human emotion through the perception system is over a predefined threshold value, its activation intensity increases.
- To maintain its well-being: The third drive is related to the robot's maintenance of its well-being with regard to psychological and physical fatigue; the first case is that when the robot have extreme anger or sadness, it stops interacting with the human and the second is that when its battery is too low to act any more, it takes a rest to recharge its battery. The psychological and physical condition is expressed as 'robot energy' valuable. If robot energy is too low, its activation intensity increases.

6. Memory System

Our system presented the emotional memories required for more natural and intelligent affective interaction with a human. Considering our research objectives, we have implemented two memory groups. One is user memory, which represents memory on previously interacted user. The other is topic memory, which represents a robot's knowledge on specific topic. When the memory cell is created, the robot's cur-rent emotional state is attached to the memory cell as emotional tag. Later, memory cells can be activated either by perception system or by emotional state of the robot. Even though many memory cells can be activated by perception system or motivation system, only one memory cell is chosen among its memory group to interact with the other subsystems. The selected memory cell is sent to the behavior system, and behavior system changes the behavior accordingly. Further, its emotional tag is also sent to the motivation system and this tag is used for determining the robot's emotion.

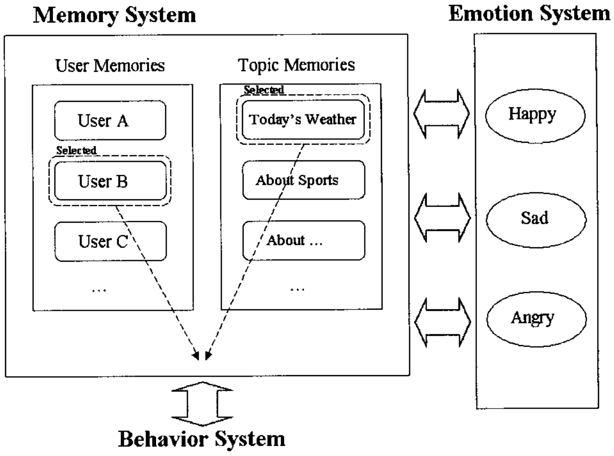


Figure 3 The activation of memory cells

7. Behavior System

The behavior system organizes its goal into a cohesive structure. The structure has three levels and branches that address the three drives: the drive to interact with a human, the drive to ingratiate itself with a human and the drive to take a task. Each branch has multiple levels, with three layers being the maximum. As the system moves down a level, the specific behavior is determined according to the affective relationship between the robot and human. As referred in our previous project [3], the structure of behavior system has three levels that address

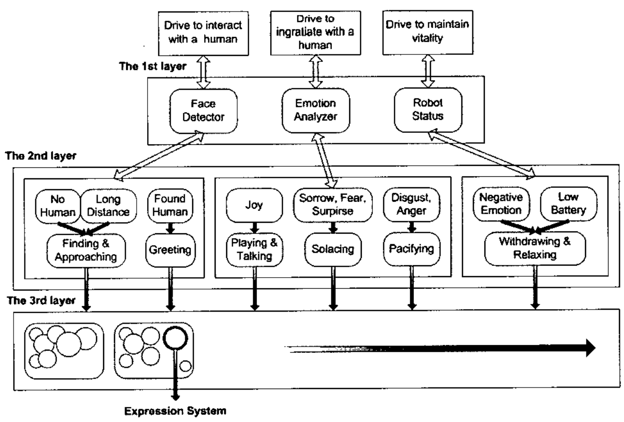


Figure 4 The behavior system hierarchy

the three drives mentioned in the motivation system. As the system moves down a level, more specific behavior is determined according to the affective relationship between the robot and human.

8. Expression System

The expression system comprises three sub systems: dialogue expression system, 3D facial expression system, and gesture expression system.

The expression system plays two important functions; the first is to execute a behavior received from the behavior system. A behavior consists of a dialog of the robot and a human. The second is to express robot's emotion. The robot expresses its own emotion through facial expressions.

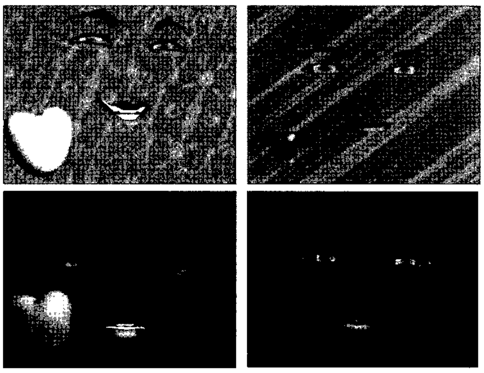


Figure 5 Facial Expressions of the robot

Dialogue is a joint process of communication sharing of information(data, symbols, context) between two or more parties. We consider three primary types of dialogue: low level(prelinguistic), non verbal, and verbal language. Among them, the robot communicates with a human through daily verbal language with appropriate gestures.

However, it is difficult to enable a robot to engage in natural dialogue with a human because of the limitation of the current technique on speech recognition, natural language processing, etc. Accordingly, we predefined dialogue flow and topics. In order to make natural dialog possible in the limit that the robot could recognize only the limited number of speech, we constructed dialog as follows: First, the robot lead actively dialog by asking user's intention ahead of him to avoid the possibility that the robot cannot understand human speech. Second, the robot answers the most frequently used responses when they cannot understand to avoid unnatural dialog.

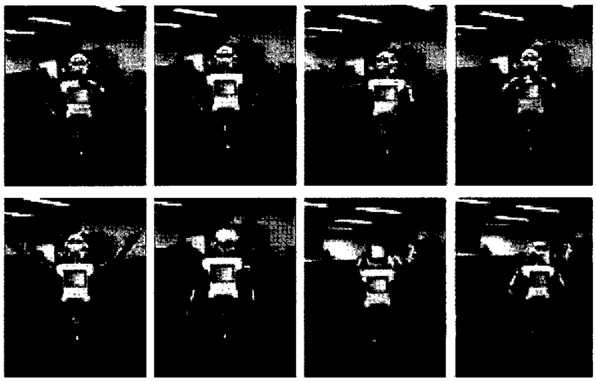


Figure 6 Gesture Expressions of AMI

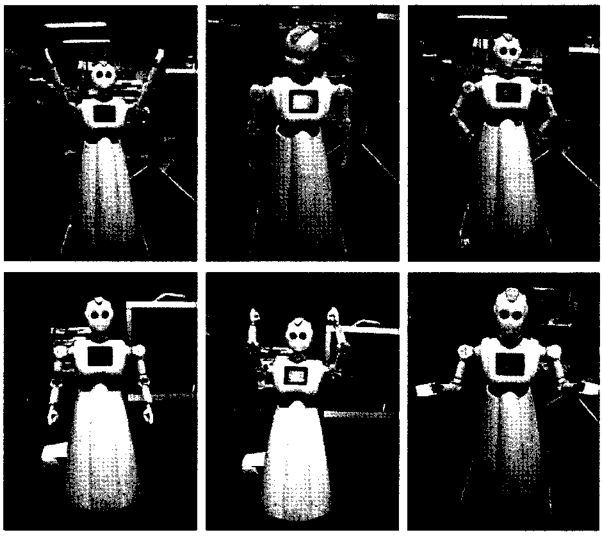


Figure 7 Gesture Expressions of AMIET

The dialogue expressions comprise the most commonly used speeches according to the selected behavior of Finding & Approaching, Greeting, Talking, Playing, Consoling, Pacifying, Withdrawing & Resting behavior group. In Finding & Approaching, the robot mainly finds a human by calling a human. In greeting, the robot says hello to a human and asks his name, etc. In Talking, dialog consists of various common topics such as hobby, weather, movies, etc. In playing, the robot plays with a human through a kind of jokes, OX quiz and Nonsense Quiz. In Consoling and Pacifying, the robot asks what the human is angry about and then makes a joke to console him or give pleasure to him. Further, the robot asks his worries by recognizing his sad emotion and listens to his sayings as his friend and counselor.

Figure 8 AMIET interacts with a human user

9. Conclusion

We showed an affective interaction system that is designed for a robot to lead interactions by recognizing human emotional status and expressing its emotion through multimodal emotion channels like a human, and behaves appropriately in response to human emotions. With explicit emotional memories on users and topics, in our system, we successfully improved the affective interaction between humans and robots. Previous sociable robots either ignored emotional memories or maintained them implicitly.

References

[1] P. Ekman and W. V. Friesen, "Facial Action Coding System: Investigator's Guide," Consulting Psychologists Press, Palo Alto, CA, 1978.

[2] Naoko Tosa and Ryohei Nakatsu, "Life-like Communication Agent - Emotion Sensing Character 'MIC' & Feeling Session Character 'MUSE," ICMCS, 1996.

[3] Hyun S. Yang, Yong-Ho Seo, Il-Woong Jeong, Hye-Won Jung, "Affective communication system with multimodality and Its Application to a Humanoid Robot," VSMM 2004.

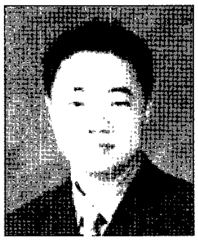
양 현 승

1976 서울대학교 전자공학과(학사)
1983 Purdue University(석사)
1986 Purdue University(박사)
2006 현재 한국과학기술원 전산학과 정교수
관심분야 : 컴퓨터 시각, 로보틱스, 인공지능, 가상현실, 유비쿼터스 컴퓨팅
E-mail : hsyang@paradise.kaist.ac.kr

서 용 호

1999 한국과학기술원 전산학과(학사)
2001 한국과학기술원 전산학과(석사)
2006 현재 한국과학기술원 전산학과 박사과정
관심분야 : 로보틱스, 인간로봇 상호작용, 로봇 시각, 웨어러블 컴퓨팅
E-mail : yhseo@paradise.kaist.ac.kr

정 일 웅



2002 아주대학교 정보및컴퓨터공학부(학사)
2004 한국과학기술원 전산학과(석사)
2006 현재 한국과학기술원 전산학과 박사과정
관심분야 : 로보틱스, 인간로봇 상호작용
E-mail : woong@paradise.kaist.ac.kr