Type of the Paper (Article)

A Contextual Emotion Appraisal System Based on a Sentential Cognitive System for Robots

Hyunsik Ahn [1]*, Sung Park[2]

[1] Department of Robot System Engineering, Tongmyong University;
[2] School of Design, Savannah College of Art and Design; spark@scad.edu
* Correspondence: hsahn@tu.ac.kr; Tel.: +82-10-2518-4777

Abstract: Emotion plays a powerful role in humans' interaction with robots. In order to express more human-friendly emotions, robots need the capability of contextual appraisal that expresses the emotional relevance of various targets in the spatiotemporal situation. In this paper, an emotional appraisal system is proposed in this study to cope with such contexts. Specifically, the Ortony, Clore, and Collins model is abstracted and simplified to approximate an emotional appraisal model in the form of a sentence-based cognitive system. The contextual emotion appraisal is modeled by formulating the emotional relationships among multiple targets and the emotional transition with events and time passing. To verify the proposed robotic system's feasibility, simulations were conducted for scenarios where it emotional interacts with humans manipulating liked or disliked objects on a table. This experiment demonstrated that the robot's emotion can variously change over time like human by using a proposed formula for emotional valence, which is moderated by emotion appraisal of occurring events.

Keywords: emotion appraisal system; contextual emotion appraisal; cognitive robotics; sentential cognitive system; HRI

1. Introduction

As the technology of Human Robot Interaction (HRI) matures, the idea of robots being companions of humans in their daily lives is becoming popular. Emotional exchange with robots facilitates a human's empathy, engagement, and collaboration. An affective robot involves understanding human emotion and expressing believable emotion to humans where the latter need to define the robot's emotional states and their expressions. Studies have focused on how emotions should be expressed on a robot's face [1, 2] and in its voice [3] and how emotions should be appraised based on recognized information or stimulation [4, 5]. One of the most significant studies on robots simulating emotion involves MIT's social robot Kismet. Its cognitive architecture was influenced by Russell's perceptual approach to emotion [6], which included critical emotional constructs, such as arousal (strong versus weak) and valence (positive versus negative) associated with the robot's internal drive (e.g., social needs, fatigue). Kismet can express affect through face, voice, and posture based on an emotional internal representation form by the affective appraisal of external stimuli, such as threatening motion or praising a human's speech.

Emotion is intertwined with context. Specifically, emotion is associated with an object or an individual at a given time. How contexts from past to present are induced to appraise a robot's emotional state is an important issue. A significant amount of research has been conducted in cognitive psychology on this subject. Ortony et al. claimed that emotion is determined by the interpretation of the context that induces emotions as reactions regarding interest toward an event or a subject. They categorized the emotional appraisal process based on 22 contexts [7].

Attempts have been made to apply psychological emotional models to robots. A study appraised emotions by defining a list of objects and actions and analyzed their relationship [4]. Another study appraised emotions using an explanatory process for recognized information [8]. However, it is difficult to consider these studies as full contextual emotion appraisals, because they appraise emotions based on objects or events at points in time, rather than in a context extending from past to present.

In this paper, an emotion appraisal system based on cognitive context of robots is proposed. For computational modeling, a simplified Ortony, Clore, and Collins (OCC) model is adopted and that is applied to a Sentential Cognitive System (SCS) to combine emotional valence with cognitive information. The state of human emotions decreases over time to neutral and some emotions are transferred to different emotions. In this paper, the change in intensity of emotional valence and the transition of emotion are modeled and formulated. For contextual emotion appraisal, the targets of emotion are divided with objects, agents, and events that make it possible to mediate objects to evaluate the emotion of the agents dealing with them. An implementation of the proposed approach is conducted with scenarios where a robot interacts with humans manipulating objects on a table.

The contributions of this paper are as follows: (1) a contextual emotion appraisal model abstracting and simplifying OCC model is implemented in a SCS of a cognitive robot. (2) the emotion appraisal system has the principal contextual characteristics by formulating the emotional relationships among multiple targets and emotional transition with events and time passing.

This paper is organized as follows. In section 2, related work on emotion appraisal studies is described. Section 3 details the contextual emotion appraisal model. Section 4 depicts a SCS based contextual emotion appraisal system. Section 5 provides the implementation of the proposed approach to a service robot and Section 6 depicts conclusions and future work.

2. Related Work

Studies have conducted to make robots understand and express emotions in various ranges. In particular, as research on robot companions to interact with each other as robots share their daily lives with humans, research to imitate human emotions has conducted for robots to express emotions as partners.

For emotional HRI, robots need a function to recognize human emotions to interact with robots. It also requires the robot to evaluate its emotions by itself from outside stimuli or information to express its state of emotion. In particular, the evaluation of emotions is not just a stimulus but a more human-friendly emotional exchange only when the robot is evaluated in the temporal, spatial, and eventful context of human beings.

Research to understand human emotions by robots is to digitize most of them. Studies to recognize emotions from the human face range from imaging signal processing to biological signals. Among the studies of human emotion recognition, Ikeda et al. used brain waves or heartbeat signals to recognize human emotions [15]. Lincon et al. provide nonverbal communication, such as recognizing emotions and displaying emotions that can identify humans to express empathy [16]. Cid et al. used in the form of a person's face to recognize emotions and recognize emotions in a speech to enable multi-modal recognition of emotions [19].

In the field of emotion appraisal of the robot itself, there are multiple approaches according to the emotion model, the data types of emotion, and the experience from the past to the present. In the view of emotion data type, the previous studies used stimuli, energies, sentences, targets, or feelings. For the case of stimuli input, Uriel et al. suggested a way to have the robot respond to emotional expressions by touch. Depending on the duration and pressure of the touch, a Bayes' rule allows the robot to evaluate and respond to four different emotions [18]. Kim et al. [24] suggested a computational reactive emotion generation model responding to temporally changing responding to stimuli using motivation, habituation, and conditional models.

In the energy model for emotion data, Lee et al. [26] distributed pleasure and arousal in the two-dimensional plane to indicate how emotions distributed with the concepts of energy, entropy, and constancy for emotional generation. Hakamata et al. conducted a research to create emotional

expressions in interactive sentences [20]. In this study, it obtained the emotional weighting vector of the word in conversation and tagged each sentence of the conversation with it as metal status. However, it does not include the maturity and transition of emotions that make a target a medium.

Jitviriya et al. used a self-organizing map to express emotions depending on the object. The motivation was determined according to objects such as color, shape, and distance [21]. However, this method also does not represent the context in which objects are used.   Samani et al. [23] designed a high level of emotional bonds between humans and robots by applying human's affection to robots. The system's advanced artificial intelligence includes three modules: Probabilistic Love Assembly (PLA) based on psychology of love; Artificial Endocrine System (AES) based on physiology of love; and Affective State Transition (AST) based on emotion.

Another approach is combining emotions with sentences. Park et al. [25] suggested that the robot uses a set of multi-model motion described as a combination of sentence type and emotion to express its behavior. Hakamata et al. suggested a model to create emotional expressions in interactive sentences [20]. In this study, it obtained the emotional weighting vector of the word in conversation and tagged each sentence of the conversation with it as mental status. However, these approaches don't include cover the maturity and transition of emotions that make a target a medium.

There are some approaches adopting a kind of context using the experience of past. Zhang et al. [17] proposed an emotional model for robots, using HMM techniques to enable past experiences to influence current emotions. This method assessed feelings based on their own emotional history, user impact, and task. Kinoshita et al. [22] proposed an emotion generation model dividing the robot's emotions with the degree of favorability as the short-term impression of a user, and degree of intimacy as the long-term impression, and expressed the accumulated affinities. The proposed model showed that robots respond differently to each user after communicating with multiple users. Kirby et al. [27] presented a generative model of affect that accounts for emotions, moods, and attitudes, including interactions between them. The model attempts to mimic the behavior, particularly with regard to long-term affective human responses. Itoh et al. [28] divided the robot's affective state into mood and emotion, which connected with conversation between a human and the robot using a dialogist.

In summary, most previous studies have an emotional evaluation as an emotion that responds spontaneously to different kinds of stimuli. The work of contextual appraisal using past experiences is a form of using accumulated emotional values over time or tagging on each dialogue sentence, and there is still a lack of research to associate contextual events related to objects and agents with temporal and spatial contexts.

3. Contextual Emotion Appraisal Model

For robots and humans to communicate by showing emotion, a theory of emotion must be interpreted and modeled from the perspective of a robot's intelligence. In general, the popular emotion model used in robot emotion research consists of information sensing, emotion appraisal, emotion generation with a personality model, and emotion expression processes. Cognitive information sensed from the external environment is analyzed to extract emotional meanings (i.e., emotion appraisal). Such appraisal is combined with a personality model, by which reactions (e.g., calm, drastic) are mediated by the type of personality. This emotion is expressed through the robot's verbal (linguistic form) or nonverbal (facial expression, gestures, postures, etc.) channels. For more human friendly emotional interaction, appraisal of emotion needs to be determined in a spatial-temporal context. In this study, such appraisal of contextual emotion is made possible by evaluating emotions on the each event that a robot experienced with the objects and agents using a cognitive system.

The Ortony, Clore, and Collins (OCC) model of emotion contains semantics regarding the appraisal of contextual emotion, as shown in Figure 1 [7]. The OCC model classifies emotion into 22 types based on the context regarding the aspects of the object, action of the agent, and consequences of the event. The outcome of the agent's action or event is elicited from contextual relationships between past and present, not merely from what is sensed. In the proposed emotion appraisal model

of this paper, the cognitive appraisal of emotion, the classification of targets of emotion, and the transition process of emotion are defined based on a simplified OCC model.
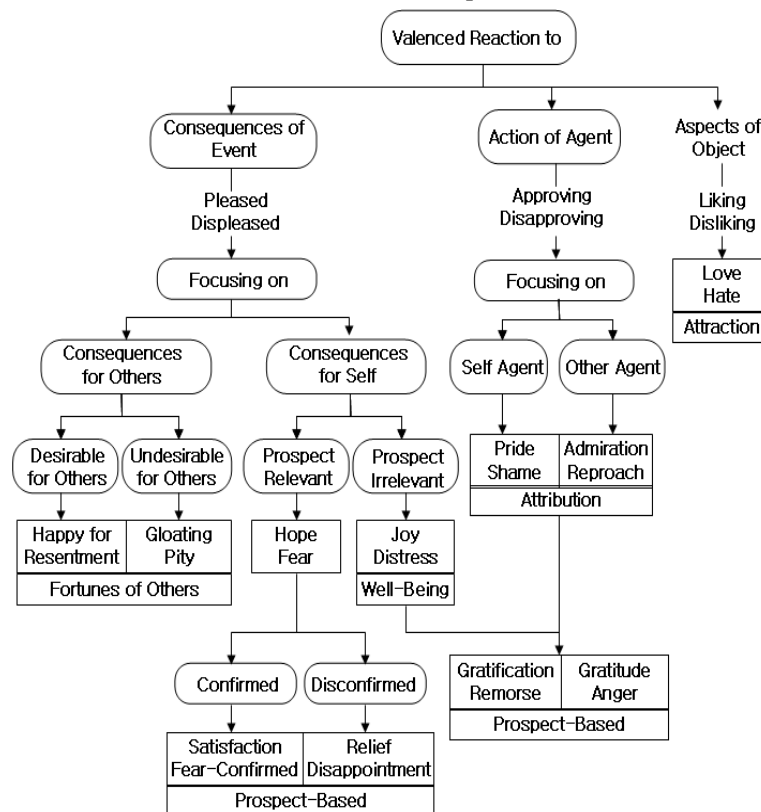
Figure 1. OCC model for emotion appraisal [7].

3.1. Emotion appraisal model by simplifying OCC model

To establish an emotion appraisal model applicable to robots from the OCC model, the following conditions are imposed:

- Primacy condition: Although the semantics of emotion can be interpreted variously, the primary meaning of the emotion is used.
- Valence condition: Rather than classifying all aspects of emotions, they are represented with the valences of primary emotional states.
- Self-centered condition: Only emotions focused on the robot itself are modeled.
- Cognitive condition: Only emotions that can be analyzed by the robot's sensory and behavioral information are used.

In this paper, the OCC model is reconstructed with these imposed conditions for being applied to robot's emotion appraisal. Because the OCC model used a dichotomy of positive and negative emotions, the 22 types of emotion are defined from 11 emotional pairs (see Figure 1). The simplified OCC model adopts 10 types of emotions with five emotional pairs (see Table 1 for the emotions) according to the primacy, valence, and self-centered conditions. For example, because Admiration–Reproach focuses on the other agent and not the robot, this pair is excluded. Gratification–Remorse is integrated with Joy–Distress because the latter is the primary emotion. As an emotional state that might arise in the future, Hope–Fear is defined as an emotion that encompasses Satisfaction–Fear-Confirmed and Relief–Disappointment because of the primacy condition, and the subtle differences would be expressed through valence. With this simplification, 10 types of emotion are classified that can be applied to robots.

Table 1. Types and contexts of emotion.

#	Emotion	Appraisal context	Dichotomy
1	Love	Aspects of Target	Positive
2	Hate		Negative
3	Pride	Action of Self	Positive
4	Shame		Negative
5	Gratitude	Action of Agent	Positive
6	Anger		Negative
7	Joy	Consequence of Event/Prospect irrelevant	Positive
8	Distress		Negative
9	Hope	Consequence of Event/Prospect relevant	Positive
10	Fear		Negative

For a robot to appraise the 10 types of emotion on its own, the appraisal process must specifically involve the events that occur with targets in the given environment. In this paper, a linguistic approach is adopted to model the structure of the emotion appraisal to be related with the targets; objects, agents, and events. To define the emotional context among the target, sentences that represent the emotions are composed using emotional verbs according to the classification, and the functions of the arguments are analyzed to define the targets of the emotions (see Table 2). Then, wordings conveying emotion are substituted with more general and primary verbs — for example, Thank with Gratitude, Rejoice with Joy, and Boast with Pride. To analyze the sentences composed with emotional verbs (Table 2) in phrase units, syntactic parsing was performed as follows using Penn Treebank tag groups [9].

Table 2. The emotional classes, sentences constructed with emotional verbs, and the targets of emotion.

#	Emotions	Sentences	Targets of emotions
1	Love	(S (NP I) (VP love (NP the object))), (S (NP I) (VP love (NP the agent) (PP by (NP the event))))	(NP the object), (NP the agent)
2	Hate	(S (NP I) (VP hate (NP the object))), (S (NP I) (VP hate (NP the agent) (PP by (NP the event))))	
3	Pride	(S (NP I) (VP boast (NP the event)))	(NP the event)
4	Shame	(S (NP I) (VP am (VP shamed (PP for (NP the event)))))	
5	Gratitude	(S (NP I) (VP thank (NP the agent) (PP for (NP the event))))	(NP the agent)
6	Anger	(S (NP I) (VP anger to (NP the agent) (PP by (NP the event))))	
7	Joy	(S (NP I) (VP rejoice (PP at (NP the event))))	(NP the event)
8	Distress	(S (NP I) (VP am (VP distressed (PP by (NP the event)))))	
9	Hope	(S (NP I) (VP hope (PP for (NP the event))))	(NP the event)
10	Fear	(S (NP I) (VP fear (NP the event)))	

S: sentence, NP: Noun phrase, VP: Verb phrase, PP: Prepositional phrase.

Syntactic parsing divides a sentence into multiple phrases. The Penn Treebank parser encompasses verb, noun, and prepositional phrases. Because of research in cognitive linguistics, it is generally accepted that the verb determines the characteristics of the sentence [10]. Therefore, the primary meaning of a sentence is determined by a verb's semantic structure, and a syntactically parsed sentence can be analyzed as an argument structure that defines the characteristics of the sentence. Once the argument structure is determined in association with the emotional verb, the

target of emotion contained in the sentence can be defined. In a sentence describing an event, the robot is the subject of the emotion and is depicted as "I." The direct target of emotion is defined in the form of a noun phrase.

Table 3 shows the modeled relationship between the emotions and targets; where circles mean direct relationship, triangles as indirective relationship, and vacant blocks as irrelevance. An object is a passive target that cannot change the environment on its own, and an agent is an active target that can change the environment, such as a human or a robot. An event is defined as something that has occurred involving a robot, an agent, or an object. The target of Love–Hate (LH) can be an object or an agent. Because Pride–Shame (PS) is an emotion toward oneself, it is defined as an emotion about an event related to oneself; therefore, other agents cannot be the target of the emotion. The target of Gratitude–Anger (GA) can be an agent or an event. Present and future events can be targets of Joy–Distress (JD) and Hope–Fear (HF), respectively.

The emotion appraisal has two kinds of sources, *a priori* and contextual sources as shown in Table 3. In the case of LH targeting objects, a robot needs to evaluate predefined emotional liked or disliked valence of the objects, because the robot don't have the feeling of satisfaction and preference to the objects. For example, because a robot cannot taste an apple, the emotional valence of the robot to an apple needs to have *a priori* definition. On the other hand, the other cases in the 10 emotions targeting agents and events need evaluate the emotional valences contextually. If a robot feels Love–Hate toward an agent, the emotion does not abruptly arise, but it is developed over time from the experience with the behavioral events of the target.

In this approach, all the emotional appraisal are computed with emotional valences with the span of -1.0 to +1.0. The negative emotion of emotional dichotomy can be extended to -1.0 according to the intensity of the emotion, and the positive to +1.0. At a certain time, the robot's emotional state is the combinational summation of emotions towards targets, agents, and events. To a certain agent at a point, the robot has the emotional state with the form of emotional valences of LH and GA as shown in the vertical block of the agent of targets in Table 3.

Table 3. The contextual emotion appraisal model with targets and source of emotions.

#	Emotions	Targets			Sources of appraisal		Emotional valence
		Object	Agent	Event	A priori	Contextual	
1	Love-Hate (LH)	○	○	△	○	○	-1.0 ~ 1.0
2	Pride-Shame (PS)			○		○	-1.0 ~ 1.0
3	Gratitude-Anger (GA)		○	△		○	-1.0 ~ 1.0
4	Joy-Distress (JD)			○		○	-1.0 ~ 1.0
5	Hope-Fear (HF)			○		○	-1.0 ~ 1.0

When an emotion is aroused with an event, the state of the emotion sustains for a while and decreases gradually with time passing. In this paper, an emotional intensity over time, $E(t)$, is defined as (1).

$$E(t) = (1 - s) \exp\left(\frac{-t^2}{d}\right) + s - ct \tag{1}$$

where d is the strong intensity duration, s is the long time sustainability, and c is diminution of long time span of emotional valence. Figure 2 shows a typical distribution graph of emotion valence with the time axis with a set of parameters.
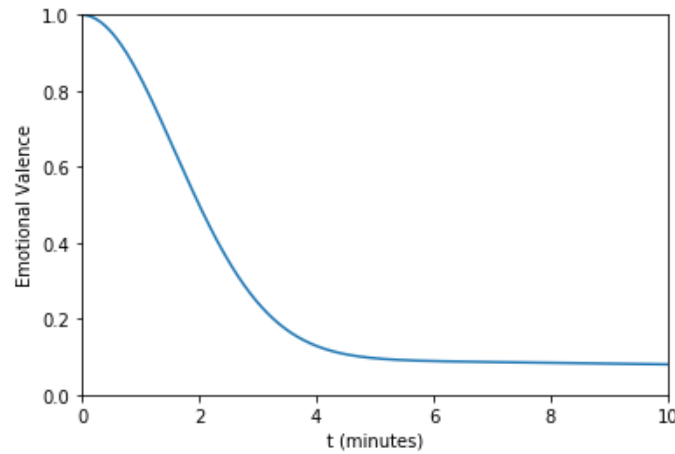
Figure 2. The temporal diminution of emotional valence with the parameters (*s*: 0.1, *c*: 0.002, *d*: 5).

In the table 3, there are the valences of the emotion pairs distributed within -1.0 to +1.0 according to the positive and negative of the emotional states. It means if a positive emotion happens, the negative emotion can cancel each other. Therefore, the valence of an emotion pair is defined at a certain time is like (2).

$$V(t) = a\,E_p(t) + b\,E_n(t) \tag{2}$$

where $E_p(t)$ is the positive emotion and $E_n(t)$ is negative emotion of the pair, and a and b are emotional combination rate in the span of 0 to 1, when the first emotion aroused due to an event.

3.2. Emotional transition model

In the proposed emotion appraisal model, the emotional state of a robot is expressed with a combinational set of multiple emotions to targets. In the case of object, the emotional state is described with the valence of LH. However, the emotional state to an agent can be described with a combination of LH and GA, and to an event with PS, JD, and HF. For example, love and gratitude can be felt at the same time, which can be understood as an emotional state being caused by the diversity in the interpretation of the context. Given this, an emotion transition model is proposed to model the emotions' multiplicity and transitional characteristics.

Figure 3 shows the emotion transition model. This model is developed with the types of emotion used in this research and is largely consistent with the mental state transition model proposed by Ren [11]. This model is supported by empirical data from human subjects. Arrow ① indicates that when an agent causing JD events to a robot is recognized, the robot may display GA toward that agent. Arrow ② indicates that LH toward an agent may rise from the GA toward that agent. JD or HF is also produced when a target of LH of objects or agents is identified (③, ⑤) or when such an encounter is expected (④, ⑥). Arrow ⑦ indicates that HF is generated when JD is expected. Arrows ⑧ and ⑨ indicate that JD and HF are produced when PS occurs or is expected to occur, respectively.

The emotion transition model explains how a single event may cause several emotional states based on the relationships among emotions.

$$V_r(t) = V_r(t-1) + \sum_{k=1}^{n} h_k V_k(t) \quad \text{when } V_k(t) > T_k \tag{3}$$

where $V_r(t)$ is a recipient and $V_k(t)$ is a sender of the emotional valence, h_k is weight factor of transition with the span of 0 to 1. Each $V_k(t)$ has a threshold value of transition (T_k). T_k is used as a level of emotional intensity, which is needed to be over for the transition.
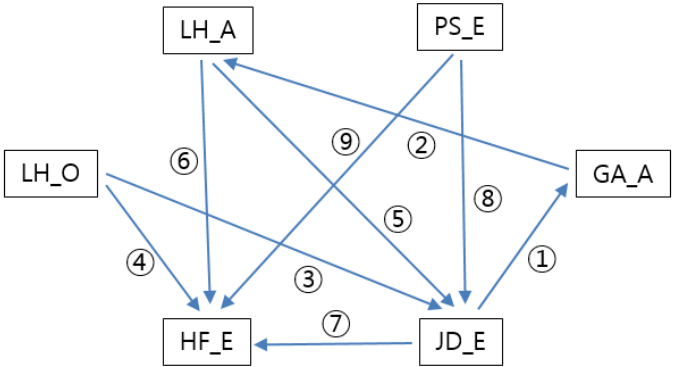
Figure 3. Transition of emotions and their mutual relationships among Love-Hate_Object (LH_O), Love-Hate_Agent (LH_A), Gratitude-Anger_Agent (GA_A), Pride-Shame_Event (PS_E), Joy-Distress_Event (JD_E), and Hope-Fear_Event (HF_E): ① An agent causing JD events, ② LH toward an agent may rise from the GA, ③ target of LH of objects is identified, ④ target of LH of objects is expected, ⑤ target of LH of agents is identified, ⑥ target of LH of agents is expected, and ⑦ HF is generated when JD is expected, ⑧ JD is produced when PS occurs, and ⑨ HF is produced when PS is expected to occur.

4. An Emotion Appraisal System on a Sentential Cognitive System

In this section, the SCS that bridges the cognitive information acquired from the sensing system to the emotion appraisal system is described.
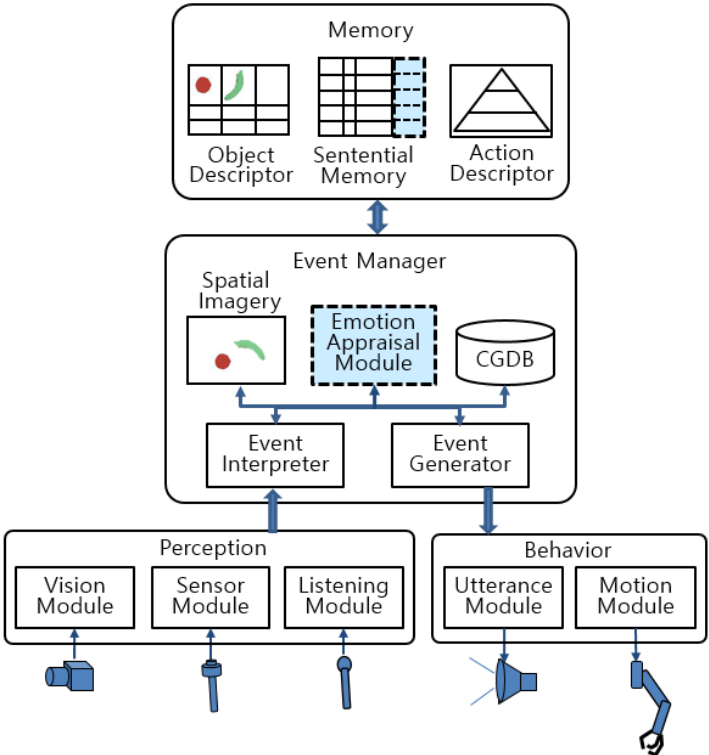
Figure 4. Schematic of an emotion appraisal system based on SCS that produces modular sentences with emotion valences from all events and stores them in the sentential and auxiliary memories.

4.1. An SCS for emotion appraisal

An emotion appraisal system based on SCS is shown in Figure 4. The SCS consists of 4 functional blocks; a memory, an event manager, a perception, and a behavior [13]. To be different from previous

SCS, it has an emotion appraisal module (dotted line) in the event manager to appraise emotion in context. One of the merit of the proposed model is that the emotion appraisal is fulfilled based on the information of the SCS. The other advantage is that an emotion can be generated with context of cognitive information. In the SCS, an event is defined as newly recognized cognitive information different from the current state in the perception and behavior modules. The visual module, listening module, and sensing module receive information from the external environment. The utterance module and motion module exert action toward the environment. In the event manager, there are an event interpreter to represent an event to a sentence and to appraise the emotion of it, an event generator to produce such behaviors as utterance and motion, and schematic imagery to allocate objects virtually for spatial inference.

With the events happened to the robot, the cognitive information is interpreted as a sentence where the valence of emotion are tagged in the sentence. The memory of the system consists of the sentential memory, the object descriptor, and the action descriptor. The sentential memory stores interpreted sentences and their valance of emotion in chronological order (dotted line in the sentential memory). Auxiliary storage modules are used to represent the events effectively. The object descriptor stores the appearances of objects and agents with their cognitive properties and valences of emotion. The action descriptor stores behavioral information linking a verbal motion to a motion program function.

4.2. Contextual emotion appraisal based on the SCS

In this section, the process by which a robot recognizes an event and appraises emotions contextually by the emotion appraisal algorithm of the cognitive system is described. The proposed contextual emotion model consists of emotion pairs according to the context shown in Table 3. Cases appraised with positive and negative emotions are defined as contrasting interpretations in the same context. The valence of emotion is quantified with real numbers between -1.0 and +1.0; a positive number denotes a positive emotion, 0 denotes a neutral emotion, and a negative number denotes a negative emotion. Emotions are appraised according to their classification as follows.

- Love–Hate (LH)

L-H toward an object is developed from the likes/dislikes information defined a priori. Love develops when a liked sensation is detected by the sensor or when a liked object or agent is recognized. The emotion of Love can also develop contextually toward an agent. For example, when an agent gives an object that the robot likes, Gratitude develops which leads to an emotional transition to Love.

- Pride–Shame (PS)

P-S is a contextually developed emotional state toward the robot itself. For example, a robot develops pride when it succeeds in completing an intended action, and shame when it fails.

- Gratitude–Anger (GA)

G-A is produced by an agent executing a certain event. Gratitude is developed from a context in which an agent's action provides something that the robot likes.

- Joy–Distress (JD)

J-D is defined as an emotion produced when the consequence of an event turns out to be favorable or unfavorable for the robot.

- Hope–Fear (HF)

H-F is defined as an emotion produced when the future consequence of an event is expected to be favorable or unfavorable for the robot.

Table 4 shows an event and emotional valence. In the vision module, an object (O_i) appeared, and a sentence was created as well as the emotional valence of the agent (A_i) who caused the emotional valence of the event of the object. The objects have an *a priori* emotional valence of LH (l). The agents have the values of emotions LH (l) and GA (g), and the events have the values of emotions of PS (p), JD (j), and HF (h). When a vision event occurs, the vision module interprets the event and generates a sentence (S_l) with the cognitive information of the object. The interpreter produces a sentence with the syntactic parsing that tags the cognitive information to the components of the

sentence. O_i is an object and x, y, z, θ are its position and pose. The emotional valence of each target has (l), (l, g) and (p, g, j, h) in between -1.0 and 1.0.

Table 4. Event and emotional valence. Each event occurs in a module (Mod.) and the objects, agents, and events have their own emotional valence according to (1) ~ (3).

#	Mod.	Sentences of events						Emotional valences		
		Verb	Arg_1	Arg_2	Space	Time	Adj.	O_i (LH)	A_j (LH, GA)	E_k (PS, JD, HF)
S_1	V	appeared	(NP O_i)	-	(PP at (NP x,y,z))	-	(PP by (NP A_j))	(l)	(l, g)	(p, j, h)
S_2	V	disappeared	(NP O_i)	-	-	-	(PP by (NP A_j))	(l)	(l, g)	(p, j, h)
S_3	V	appeared	(NP A_j)	-	(PP at (NP the front))	-	-	-	(l, g)	(p, j, h)
S_4	V	disappeared	(NP A_j)	-	-	-	-	-	(l, g)	(p, j, h)

5. Experiments and Results

5.1. Implementation

The proposed emotion appraisal system was implemented using a sentence-based cognitive system, and experiments were conducted using scenarios to test the feasibility of the system. The experimental scenarios involve analyzing robot's emotional state when an agent gives or takes away an object that the robot likes or dislikes. Figure 5 shows the robot used in the experiment that is embedded with the emotion appraisal system. The emotion appraisal system was implemented with an IBM PC in the Visual C++ environment and implemented by programming the modules depicted in Figure 4. Link Grammar Parser's Penn Treebank was used as the syntactic parser [12], and automatic parsing was enabled so that events could be represented in sentences and emotions related to such sentences could be appraised [13]. An RGB-D camera, Microsoft Kinect, attached at the head of the robot, was used for the vision module of SCS [14].

As explained in Table 3, emotions toward a target are classified into 10 types consisting of five pairs of contrasting meanings. Objects are appraised with LH, agents with LH and GA, and events with PS, JD, and HF. A single emotion is assigned to one event, and transitions could take place among the five types of emotional state as shown in Figure 3.

When visual events were happened, the vision module of SCS acquired RGB-D images captured by Kinect were processed to recognize the positions and poses of objects on a table. To recognize the objects in the RGB image of the input data, You Only Look Once (Yolo), a Convolution Neural Network (CNN), was adopted [29]. The vision module first found the Bounding Box (BBX) and label of the object. OpenGL libraries were used to get x, y, z coordinates of the cloud points using perspective transformation of depth data.

Figure 6 shows an RGB-D image and 3D recognition of objects in the vision module. Figure 6(a) shows object recognition results using Yolo from the RGB image, and Figure 6(b) shows the 3D view of table which was transformed from the depth data with RGB texture mapping. In the 3D view, the center position of the object is identified on the planar coordinate system. The labels and positions of the objects are stored in the object descriptor.

Figure 5. Emotion appraisal system implemented in a robot (TUBO) and an agent.
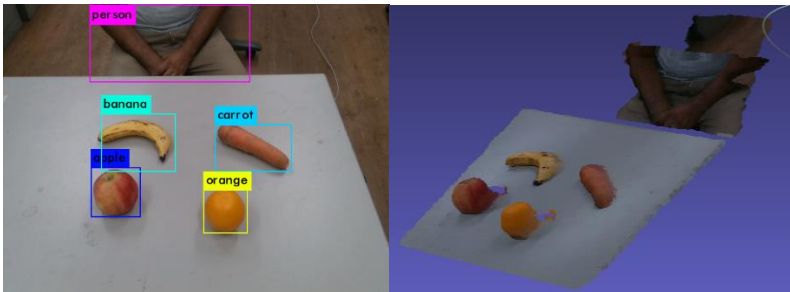
Figure 6. Example of object recognition from RGB image and 3D image transformed from depth data to the table surface plane.

5.2. Experimental results

The scenario of the experiments was testing the state of emotions when agents (human) bring or take fruits that the robot likes or dislikes on a table. When the scene of the objects are changed in the presence of an agent, the robot considers that the agent has taken an action and regard it as an event. Every event is represented as a sentence, and emotion appraisal is performed simultaneously. In the experiments, when an agent brings like or dislike fruits in front of the robot, the event affects the JD of the event and GA of the agent (Figure 3). The GA of the agent can be directly computed according to the *a priori* LH emotion of the fruit (dotted line). The GA of the agent can be transited to LH of the agent, which can also transited to the JD of the event.

The emotional valences of robot's *a priori* emotion to objects were predetermined as shown in Table 5. It assumes that the robot has positive *a priori* LH emotion to apples (0.5) and oranges (0.3), but negative to bananas (-0.2) and carrots (-0.3).

Table 6 shows the parameters of variation of emotional intensity $E(t)$ over time depicted in (1); LH of objects (LH_O), LH of agents (LH_A), GA of agents (GA_A), and JD of events (JD_E). Figure 7 shows the variation of emotional intensity over time by applying the parameters of the targets of Table 6.

Figure 8 shows the result of object recognition of the series of events obtained using YOLO in the vision module, indicating the BBX's and labels of objects. An agent was labeled as a person and each fruit was labeled and the center position of them was obtained from calculated coordinates by perspective transformation of the depth data.

Table 5. The emotional valences of a priori L-H emotion to objects in the experiments.

Objects	*A priori* LH_O Valences
apples	0.5

orange	0.3
bananas	-0.2
carrot	-0.3

Table 6. The parameters of an emotional intensity over time *E(t)* of (1); L-H of objects (LH_O), L-H of agents (LH_A), G-A of agents (GA_A), and J-D of events (JD_E) in the experiments.

Targets	Parameters of *E(t)* of (1)		
	s	*d*	*c*
LH_O	1	1	0
LH_A	0.5	8	0.002
GA_A	0.1	5	0.002
JD_E	0.2	7	0.002
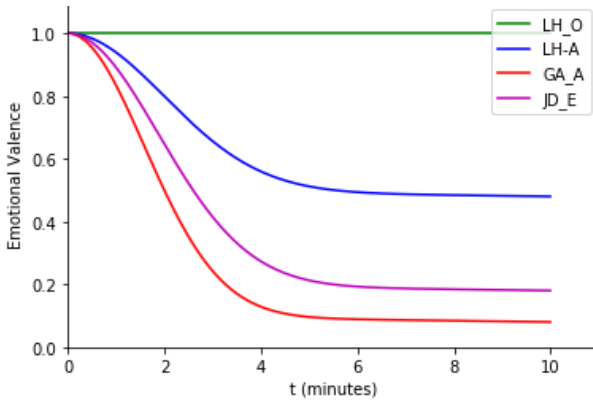
Figure 7. The variation of emotional intensity over time applying the parameters of the targets.
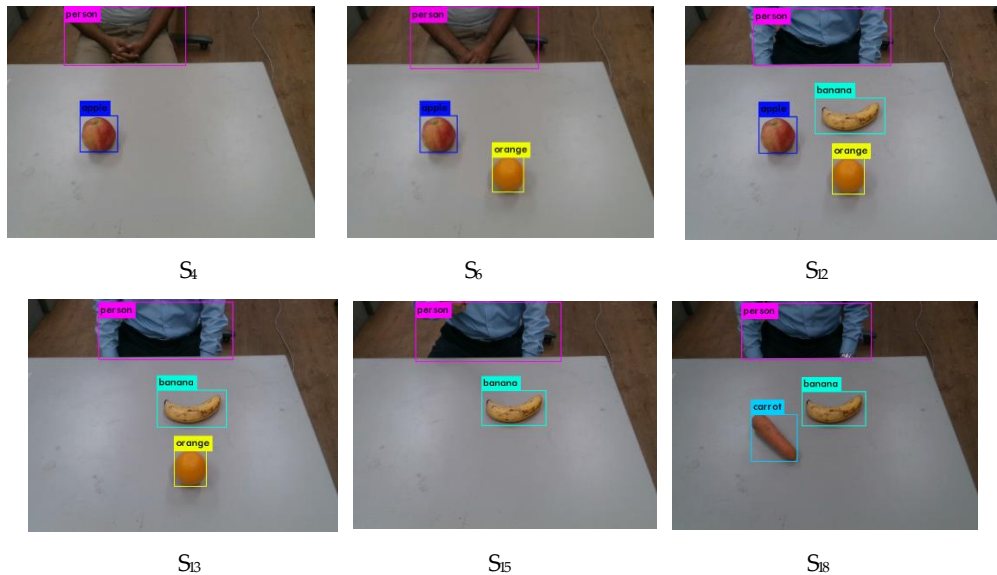
Figure 8. The object recognition results of each events (Captions are the sentence number in Table 7).

Table 7 shows the sentential memory of each event with the emotional valence which was obtained from emotional appraisal system. Events that occurred with agents and objects in front of the robot were stored in the sentential memory with chronological order. In the case of S_1–S_3 and S_9–S_{11}, new agents, A_1 and A_2, were appeared and the robot had a neutral emotional valence to them at first.

Figure 9 shows the variation of valences of emotions over time for the targets of two agents. When an apple (a priori LH valence: 0.5) appeared by agent A_1, GA-A_1 was sustained for the time being and decreased with the graphical form of GA_A in Figure 7. Because GA-A_1 was over the a threshold value (0.5) with a new event, the robot remarks "Thank you John" to the agent A_1 like (S_5, S_7). With the appearance of orange (a priori LH valence: 0.3), the valence of GA_A_1 was over the threshold value (T_k: 0.7) which makes an emotional transition to LH-A_1 with a weight factor (h_k: 0.7) of (3). Because LH_A_1 was over the threshold value (0.5) with a new event, the robot uttered "Thank you John" to the agent A_1 (S_8). When an agent A_2 brought fruits having negative a priori LH_O valences or took out the liked fruits, the GA_A_2 had negative values and expressed negative utterances such as "I am angry Tom" (S_{14}, S_{16}, S_{19}). The negative valence of GA_A_2 also made an emotional transition to LH_A_2 with the negative threshold value (T_k: -0.7) (S_{15}) and the utterance of "I hate you Tom" (S_{17}). In the Figure 9, we can see that the valence of GA decrease to neutral in short time whereas the valence of LH sustains longer by adopting the parameters of $E(t)$ in Table 6 like the graphical characteristics of each emotion as shown in Figure 7.

Figure 10 shows the variation of valences of emotions toward the all events over time for event. There were multiple events such as the appearance and disappearance of fruits (LH_Obj) and the love and hate events to agents (LH_Agent). The JD_Event of the robot is appraised by the combinational transitions of LH_Obj and LH_Agent according to (3). LH_Obj was computed by summation of the positive and negative events with the emotional combination rate (0.7) when the robot encounter the fruits and LH_Agent is also computed with the integration of positive and negative LH emotion toward agents. The JD_Event was computed with the summation of both LH_Obj and LH_Agent with the emotional combination rate (0.7). The JD_Event was positive at first but was converted to negative at the time of 10:17:32 (S_{15}).

These experiments demonstrated contextual emotion appraisal in the multiple point of view. First, the proposed contextual appraisal system showed that it was possible to evaluate emotions to the target of agent with events related to liked or disliked objects as well as just computing a certain situation of a target, which means that the proposed model can support contextual emotion appraisal. Secondly, the emotional transition model can not only make a certain emotion sustain for the time being and decrease to neutral state, but also develop it to the other emotions by combination of the current emotions. This transition characteristics of the model can be a principal aspect of contextual emotion appraisal. Due to these merits, the proposed emotion appraisal system built on a SCS of a robot can be practical for being used to human friendly emotional interaction between human and robots.

Table 7. The sentences of events and emotional valences stored in the sentential memory.

#	Time (hh:mm:ss)	Module	Sentences of events	Emotional valences of targets		
				A_1 (LH, GA)	A_2 (LH, GA)	Events (PS, JD, HF)
S_1	**10:13:02**	**V**	(S (NP A new agent) (VP appeared (PP at (NP the front)))))	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0, 0.0)
S_2	10:13:05	U	(S What (S (VP is (NP your name)))))	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0, 0.0)
S_3	10:13:09	L	(S (NP I) (VP am (NP John))))	(0.0, 0.0)	(0.0, 0.0)	(0.0, 0.0, 0.0)
S_4	10:13:25	V	(S (NP An apple) (VP appeared (PP at (NP (173, 296))))))	(0.0, 0.50)	(0.0, 0.0)	(0.0, 0.28, 0.0)
S_5	10:13:28	U	(S (VP Thank (NP you)))	(0.0, 0.50)	(0.0, 0.0)	(0.0, 0.28, 0.0)
S_6	10:14:06	V	(S (NP An orange) (VP appeared (PP at (NP (343, 162))))))	(0.53, 0.76)	(0.0, 0.0)	(0.0, 0.73, 0.0)
S_7	10:14:10	U	(S (VP Thank (NP you)))	(0.53, 0.75)	(0.0, 0.0)	(0.0, 0.73, 0.0)
S_8	10:14:14	U	(S (NP I) (VP love (NP you)))	(0.53, 0.74)	(0.0, 0.0)	(0.0, 0.72, 0.0)
S_9	10:16:03	V	(S (NP A new agent) (VP appeared (PP at (NP the front)))))	(0.31, 0.43)	(0.0, 0.0)	(0.0, 0.49, 0.0)
S_{10}	10:16:06	V	(S What (S (VP is (NP your name)))))	(0.31, 0.43)	(0.0, 0.0)	(0.0, 0.48, 0.0)
S_{11}	10:16:11	L	(S (NP I) (VP am (NP Tom))))	(0.29, 0.42)	(0.0, 0.0)	(0.0, 0.47, 0.0)
S_{12}	10:17:07	V	(S (NP A banana) (VP appeared (PP at (NP (635, 393))))))	(0.15, 0.35)	(0.0, -0.20)	(0.0, 0.24, 0.0)
S_{13}	10:17:23	V	(S (NP An apple) (VP disappeared))	(0.13, 0.33)	(0.0, -0.69)	(0.0, -0.07, 0.0)
S_{14}	10:17:26	V	(S (NP I) (VP am (ADJP angry)))	(0.12, 0.33)	(0.0, -0.70)	(0.0, -0.07, 0.0)
S_{15}	10:17:32	U	(S (NP An orange) (VP disappeared))	(0.12, 0.32)	(-0.69, -0.99)	(0.0, -0.63, 0.0)
S_{16}	10:17:35	V	(S (NP I) (VP am (ADJP angry)))	(0.11, 0.32)	(-0.69, -0.99)	(0.0, -0.64, 0.0)
S_{17}	10:17:38	U	(S (NP I) (VP hate (NP you)))	(0.11, 0.32)	(-0.69, -0.98)	(0.0, -0.64, 0.0)
S_{18}	10:19:02	V	(S (NP A carrot) (VP appeared (PP at (NP (201, 278))))))	(0.07, 0.27)	(-0.82, -0.61)	(0.0, -0.62, 0.0)
S_{19}	10:19:05	U	(S (NP I) (VP am (ADJP angry)))	(0.07, 0.27)	(-0.80, -0.61)	(0.0, -0.61, 0.0)

(V: Vision module, L: Listening module, U: Utterance module).
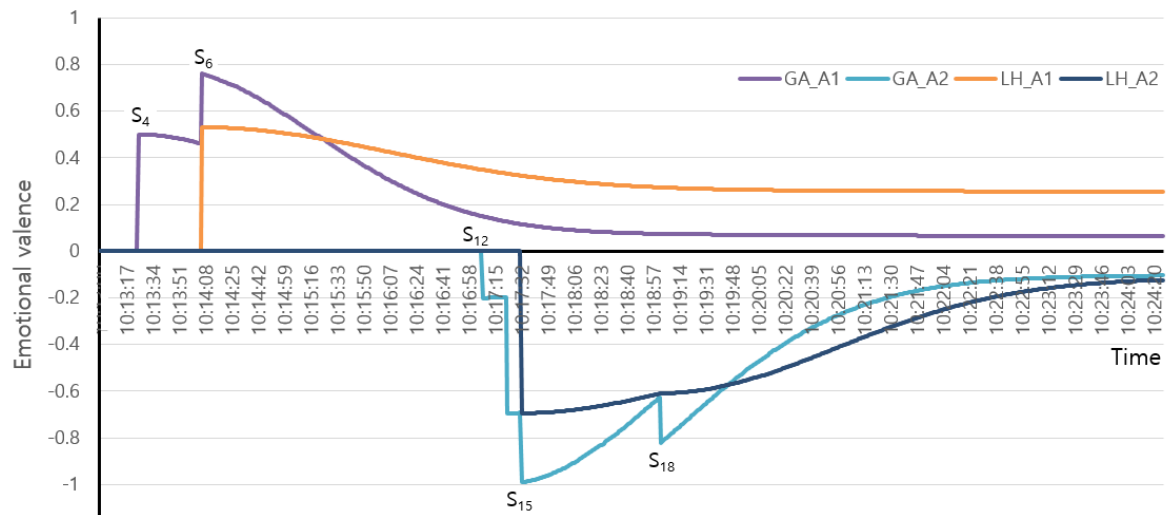
Figure 9. The variation of emotion appraisal toward the agents (A₁ and A₂).
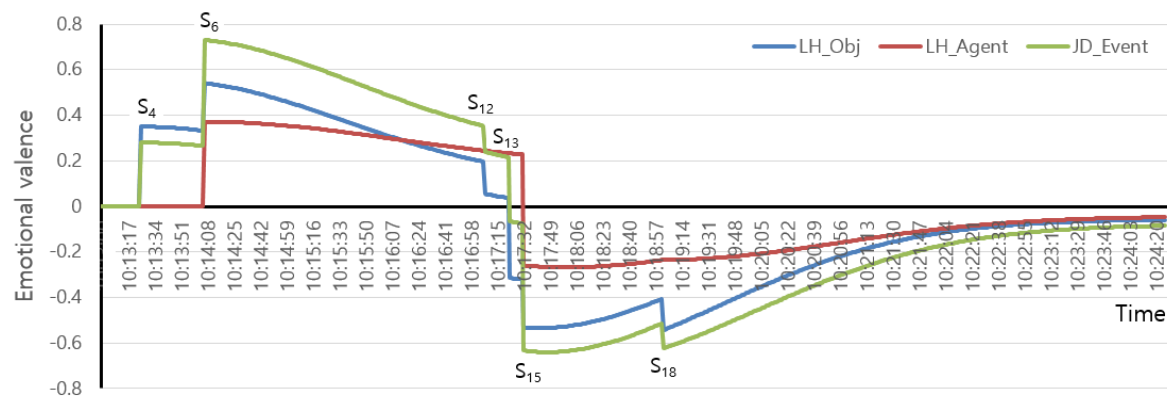
Figure 10. The variation of emotion appraisal toward the events.

6. Conclusion

A contextual emotion appraisal system using a sentence-based cognitive system that can be applied to HRI was proposed. Ten types of emotion for context-based emotion appraisal were classified by redefining the psychological OCC model. Syntactic parsing and analysis of sentences using emotion verbs showed that contextual characteristics of emotions and the targets of emotions such as object, agents, and events. The contextual appraisal was modeled by formulating the emotional relationships among different targets and the emotional transition over time. Furthermore, the emotion appraisal system was implemented using a sentence-based cognitive system. The emotional valences of targets were evaluated with the each event which causes emotional context, tagged on the parsed sentence expressing the event, and stored in the sentential memory of SCS.

The experiment demonstrated that contextual emotion appraisal can be achieved by evaluating the variation of a robot's emotion when agents give or take away an object that the robot likes or dislikes. For the proposed model to be applied to practical HRI, the cognitive capability of perception and behavior modules needs to be advanced, including 3D motion recognition and ontological classification of all the targets. Deep-learning approaches have grown in interest lately and could be adopted for the advancement of this field of research. Also, an integrated cognitive model that includes emotion appraisal from verbal and nonverbal cues could increase the validity of a human-friendly HRI.

Funding: This study was funded by Basic Science Research Program through the National Research Foundation of Korea (grant number 2012-0003116).

Acknowledgments: This research was supported by Tongmyong University Research Grants 2019 (2019F006) and Basic Science Research Program through the National Research Foundation of Korea (2012-0003116)..

Conflicts of Interest: The author declares no conflict of interest.

References

[1]   M. Hashimoto, C. Yokogawa, T. Sadoyama, Development and control of a face robot imitating human muscular structures. International Conference on Intelligent Robots and Systems (2006) 1855-1860.

[2]   H.-S. Lee, J.-W. Park, M.-J. Chung, A linear affective space model based on the facial expressions for mascot-type robots. SICE-ICASE International Joint Conference (2006) 5367-5372.

[3]   C. Breazeal, Emotive qualities in lip synchronized robot speech. Advanced Robotics 17(2) (2003) 97-113.

[4]   G.-Y. Park, S.-I. Lee, W.-Y. Kwon, J.-B. Park, Neurocognitive affective system for an emotive robot. Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robot and Systems (2006) 2595-2600.

[5]   R.C. Arkin, M. Fujita, T. Takagi, R. Hasegawa, An ethological and emotional basis for human-robot interaction. Robotics and Autonomous Systems 42(3) (2003) 191-201.

[6]   J. Russell, Evidence for a three-factor theory of emotions. Journal of Research in Personality 11(3) (1977) 273-294.

[7]   A. Ortony, G.L. Clore, A. Collins, The cognitive structure of emotion. Cambridge University Press. 1988.

[8]   S. Oh, J. Gratch, W. Woo, Explanatory style for socially interactive agents, Proceedings of 2nd International Conference on Affective Computing and Intelligent Interaction (2007) 534-545.

[9]   M.P. Marcus, B. Santorini, M.A. Marcinkiewicz, Building a large annotated corpus of English: the Penn Treebank computational linguistics 19(2) (1993) 313-330.

[10] S.M. Garnsey, N.J. Pearlmuttter, E. Myers, M.A. Lotocky, The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. Journal of Memory and Language 37 (1997) 58–93.

[11] F. Ren, Affective Information processing and recognizing human emotion. Electronic Notes in Theoretical Computer Science 225 (2009) 39-50.

[12] http://www.abisource.com/projects/link-grammar. Accessed 1st July 2018.

[13] H. Ahn, A sentential cognitive system of robots for conversational human–robot interaction, Journal of Intelligent & Fuzzy Systems 35 (2018) 6047–6059

[14] http://developer.microsoft.com/en-us/windows/kinect. Accessed 1st July 2018.

[15] Y. Ikeda, R. Horie, M. Sugaya, Estimating emotion with biological information for robot interaction, Procedia Computer Science 112 (2017)1589-1600.

[16] J.A. Rincon, et al., EmIR: an emotional intelligent robot assistant, CEUR Workshop Proceedings (2018)

[17] X. Zhang, et al., A robot emotion model with history, 2017 IEEE International Symposium on Robotics and Intelligent Sensors (2017) 230-235.

[18] U. Martinez-Hernandez, T.J. Prescott, Expressive touch: control of robot emotional expression by touch, 25th IEEE International Symposium on Robot and Human Interactive Communication (2016) 974-979.

[19] F. Cid, L.J. Manso, P. Nunez, A novel multimodal emotion recognition approach for affective human robot interaction, IEEE/RSJ International Conference on Intelligent Robots and Systems (2015) 1-9.

[20] A. Hakamata, F. Ren, S. Tsuchiya, Human emotion model based on discourse sentence for expression generation of conversation agent, 2008 International Conference on Natural Language Processing and Knowledge Engineering (2008)

[21] W. Jitviriya, E. Hayashi, Design of emotion generation model and action selection for robots using a selforganizing map, 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (2014) 1-6.

[22] S. Kinoshita, H. Takenouchi, M. Tokumaru, An emotion-generation model for a robot that reacts after considering the dialogist, 2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM) (2014) 1-6.

[23] H.A. Samani, E. Saadatian, A multidisciplinary artificial intelligence model of an affective robot, International Journal of Advanced Robotic Systems 9, (2012) 1-11.

[24] H.-R. Kim, et al., Designing reactive emotion generation model for interactive robots, The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (2010) 18-22.

[25] J.-C. Park et al., Robot's behavior expressions according to the sentence types and emotions with modification by personality, 2010 IEEE Workshop on Advanced Robotics and its Social Impacts (2010) 105-110.

[26] W.H. Lee, et al., Robot's emotion generation model for transition and diversity using energy, entropy, and homeostasis concepts, Proceedings of the 2010 IEEE International Conference on Robotics and Biomimetics (2010) 555-560.

[27] R. Kirby, et al., Affective social robots, Robotics and Autonomous Systems 58 (2010) 322–332.

[28] C. Itoh, S. Kato and H. Itoh, Mood-transition-based emotion generation model for the robot's personality, Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics (2009) 2878-2883.

[29] J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, CVPR (2017) <u>https://pjreddie.com/publications/</u>. Accessed: 2019-09-01.