Robot Emotion Representation Activated by Visual Images and Memories

Dang Thi Le Quyen

School of Information Science
Japan Advanced Institute of Science and Technology
March, 2017

Master's Thesis

Robot Emotion Representation Activated by Visual Images and Memories

1510031 Dang Thi Le Quyen

Supervisor : Professor Nak-Young Chong
Main Examiner : Professor Nak-Young Chong
Examiners : Professor Kazunori Kotani
Associate Professor Atsuo Yoshitaka

School of Information Science
Japan Advanced Institute of Science and Technology

February, 2017

Abstract

In human development process, emotions play a significant role for obtaining survival and social skills. Emotions, which are a complex phenomena of humans coupled with social context, consist of multiple different components such as cognitive processes. It helps to reflect the way humans are affected, how they adapt to different factors overtime from the environment and discriminate different personalities through the produced emotional responses. Therefore, the relationship between an individual with others are strengthened. In other words, emotions help to facilitate believable interactions among humans. Recently, emotions have been studied to model artificial emotions for robots to create natural Human-Robot Interaction (HRI). This interest is due to the fact that humans tend to treat robots by the way they treat others. Social robots can generate and express emotions via face, gesture and speech. However, social robots can perform uniform limited functionalities due to the small number of generated basic emotions. Robots then loose human attention and interest after a short period of time. In the attempt of facilitating HRI based on emotions, cognitive memory architectures have been investigated based on the role of emotions in enhancing memory performance. Robots thus are provided with a capacity to simulate human cognition in some specific cases. In addition, robots can acquire knowledge and generate emotions with social context due to the model of human-inspired social referencing. This process acts strongly in human infants, helps to form a basic novel knowledge and generate personal affective appraisal about a particular object of an individual for shaping their emotions and emotional responses to the object in the future. However, those previous works on integrating emotions with cognitive memory architectures, included human-inspired social influence, disable robots in actively acquiring knowledge and generating emotions. For one stimulus, multiple robots react similarly based on the same generated emotion for several interactions with humans despite of different prior experiences and knowledge.

In order to enable robots to actively acquire knowledge, generate human-like emotions and facilitate a believable HRI, we propose our robot emotion representation through exploring human-inspired emotion activation cues. Investigated cues include social effects and past experiences. For social effects, we investigate two main fundamental aspects which are referencing and sharing because these aspects play a significant role in humans for forming personal interpretation about the environment and generating emotions in response based on knowledge of others. We model these aspects as human guide to share knowledge and direct robot emotion generation. Robots can become a passive learner to mimic human emotions and accept human guide directly for gaining knowledge through HRI at the initial state. The personal obtained knowledge of robots is consolidated and maintained in a developmental memory architecture for future recalls. The acquired knowledge can influence robot personal emotions when being retrieved. Robots can use recalled experiences to negotiate and evaluate human guide in order to obtain a mode detailed interpretation about the environment and generate emotions in response as an active partner. Furthermore, we propose to design a developmental memory architecture which does not aim at modeling some specific human cognition as other previous cognitive memory architectures do. Our proposal helps robots to acquire knowledge continuously during interactions with humans and to adapt intelligently to the environment in general due to the relationship among memory components and the role of contextual information. Moreover, the proposed model of representing robot emotion enables robots to learn from humans and generates human-like emotions personally; thereby facilitating the interactions between humans and robots.

We designed two main experiments based on a designed scenario to evaluate our model performance. In both experiments, our robots are initialized with neutral non-arousing emotional state in which robot personal emotions are not affected by any stimulus and have no prior knowledge. During HRI, robots can get human guide to be able to extract features of the stimulus or object based on human guide in order to form personal appraisals consolidated in robot long-term memory for future retrievals and generate emotions. In order to give responses, our robots represent their emotions on the valence-arousal coordinate. The first experiment aims at exploring the model of human-inspired social effects in robot emotion generation. Robots can play the role of passive learner to mimic human emotions and to directly accept human guide for gaining knowledge. Robots can also use gained knowledge to become an active partner when negotiating with human guide to acquire past experiences for forming object appraisals and generate emotions in response to a certain stimulus in the environment. Thus, both aspects of human-inspired social effects are fully investigated in our model. In the second experiment, we observe the generated emotions of two robots for several interactions with the same human. We initialize these robots with a neutral non-arousing emotional state and empty long-term memory. One set of images which are ordered in different sequences, are given to robots. Besides, human guide is provided equally for robots. At the fourth interaction and the eighth interaction, these robots are affected by the same stimuli. However, the result shows that two robots react differently in emotion generation. The reason is that these robots gain different experiences through previous interactions. Different experiences act differently in driving robot emotion generation mechanism.

In conclusion, our proposed robot emotion representation enables robots to generate emotions and develop intellectually in general due to a cognitive memory and an exploration of human-inspired social effects. The robot cognitive memory is designed based on a developmental memory architecture. Based on this memory, robots can form and maintain object affective appraisals during HRI for future retrieval. Appraisals help robots to acquire interpretation about stimuli from the environment and gain experiences. Personal experiences which can help to discriminate a robot with other peers, are learned and shared by humans by the model of human-inspired social effects. Therefore, humans can direct robot interpretation about the environment as a teacher and a partner. Robot emotions are shaped by humans would be more favorably accepted by humans.

Keywords: Human-Robot Interaction, Robot Emotion, Developmental Architecture, Social Sharing, Social Referencing

Acknowledgment

I am greatly thankful to my supervisor, Professor Nak-Young Chong for his gentle encouragement and instruction for my Master course. He has always inspired me in my research. I could not finish this work and get any achievement without his support. It was a pleasure to work with him.

I also would like to thank Assistant Professor Francois Pierre Andre Bonnet, Assistant Professor Jeong Sungmoon, Associate Professor Nao Hirokawa of JAIST and Professor Xavier Defago of TITECH for their valuable comments and advises during the time I have been studying and doing my research in JAIST. Assistant Professor Francois Pierre Andre Bonnet and Professor Xavier Defago have kindly answered my questions during my studying while Assistant Professor Jeong Sungmoon has done the same for my research since the day I started to make initial steps in doing research. It was worthwhile for me to work with Associate Professor Nao Hirokawa in my minor research and receive expensive knowledge in a different field from my major.

A special thank to my friends who always believe in me, make me confident in myself and being with me during hard time. Nguyen Tang Tri Duc, Pham Hoang Bao Nhien and Ngo Van Thuan have brought me a lot of useful advises in daily activities and aided me to understand topics in my courses. Hosun Lee, Ferdian Adi Pratama, Alessandro Carfi and especially Abdullah Al Redwan Newaz and Kshitij Tiwari have helped me to improve my communication skills and gave me significant comments and advises for my presentations, writing and research. I also want to thank Thida Khin, Trinh Thi Thanh Mai and Lim Sin Yee due to their presence, encouragement and sharing throughout my difficult life in a foreign country. I also want to say a sincere thank to Valentin Honore who provided me with valuable guidance and helpful advises to move on when I faced to a stressful unexpected situation without having the presence of my parents in Japan. All of them make me feel like a younger sister and have another home in JAIST. It is not easy to find such great friends when living outside of my country, Vietnam, and having their support throughout a long duration. I am very grateful to be their friend.

Finally, I want to say the deepest thanks to my family for the unconditional love and continuous support. Based on the constant encouragement from my family and especially important guidance from my father, I can follow my dream of doing research and chase goals of my life. It was because of my family, I could continue my study under pressure of research and stress of the unexpected situation and finally got an academic degree in Japan. I am very thankful for their presence. I would like to consecrate my work to my family for their endless love, support and encouragement throughout my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

Emotions are known as a key factor that drive behavior cognition selection during human-human interactions to enhance belief among humans. The role of emotions has been modeled to robots for facilitating human-robot interactions. Robotic researchers modeled emotions for enabling robots to express emotions via facial expressions, music sounds, and body languages [32] [33] [29]. Other research works in cognitive science use robot emotions to give feedback to humans such as presenting robot internal states and drive robot behavior selection processes [30] through exploring the role of memory in robot development. Also, some works focus on the role of emotion in memory enhancement, thereby modeling emotions as a parameter or function of memory to enable robots to remember past experiences and use them for driving cognition processes and finishing given tasks effectively [20] [44].

However, emotions of an individual are activated and influenced by not only past experiences but also people around as the impact of social effects. Social effects include two main fundamental components: social referencing and social sharing, which helps human to form personal knowledge and generate emotions with social context in response to a certain stimulus from the environment based on the interpretation from others. On the one hand, social referencing shows strong effect to human infants for providing them with a capacity to mimic their parent emotions and form personal basic knowledge from parent guide. On the other hand, social sharing allows children and adult to combine gained knowledge with guide from others to form a more detailed interpretation about the environment and a larger number of complex emotions such as guilty or shame. The effects of social sharing and social referencing can be modeled for robots to create artificial emotions and direct robot knowledge acquirement. Social referencing can be apply to provide robots an available resource of valuable information about the world for sharpening robots knowledge as similar as the interpretation of humans while social sharing can help to enable robots to independently form self-evaluation and learn continuously from humans in respond to ambiguous stimuli from the environment or during human-robot interactions.

In conclusion, we specify our thesis problem statement as follows: Making robots more human-like to facilitate human-robot interactions is challenging when incompletely considering the role of social effects and specifically simulating human cognition at representing artificial emotions for robots.

1.2 Research Objective

To solve the identified problem, and make robot emotions be more favorably accepted by humans, we state our research statement in the following way:

It is possible to create a believable human-robot interaction via providing robots a capacity to develop intellectually and generate emotions in a human-like way based on the inspiration from human developmental and social psychology theories about cognitive memories and emotion activation cues. Different personal knowledge may form robot personalities which enable robots to have different interpretation about a certain stimuli from the environment even being influenced by the same guide from humans. Thus, two different robots can generate different appraisals about the same stimuli; thereby generating different emotions in respond to the stimuli.

1.3 Research Approach

In this research, we aimed at developing a robot emotion representation based on modeling human emotion activation cues: social effects and personal knowledge as human guides and robot past experiences respectively during human-robot interactions. Characteristics of this representation are presented as below:

- **Providing robots with capacity to conceptualize, consolidate and recall a specific experience by forming novel object affective appraisals in a developmental memory architecture**. Robots achieve human-provided contextual information to localize objects and extract visual features of those objects and form object affective appraisals. The information is integrated with a robot personal emotion to be consolidated in a developmental long-term memory architecture for future retrieval. Robots are able to describe past experiences including prior knowledge about those experiences and engaged emotions when they are recalled based on the similarity of personal past experiences and the new object affective appraisals.

- **Fully considering two fundamental aspects of social effects to enable robot abilities to learn from and share with humans**. Robots may generate emotions differently from humans after obtaining and recalling personal novel knowledge and guides for similar stimulus in the environment. For instance, robots do not have any prior knowledge at initial state similarly to human infants who are younger than 18-month old. The novel knowledge about stimuli in the environment and emotion generation process of robots are learned directly from humans through interactions. After gaining knowledge, robots can form their own affective appraisals and generate emotions based on these appraisals when seeing similar stimuli. We expect that robots can play both role of a passive learner and an active partner when getting guide from humans during human-robot interactions.

- **Generating emotions for continuously adaptation to the environment during interactions with humans based on past experiences and human guide influence**. Different past experiences can lead to generate different knowledge for the same stimulus in the environment; therefore, different reactions for adaptation process are generated. In other words, one robot can have different emotions due to different times of being affected and different emotional state at the time of encoding the stimulus. Furthermore, different robots can generate different emotions due to unequal prior knowledge and the way of interpretation stimuli from the environment or humans. This idea is similar to the Principle of Subjectivity [72] which has been applied in [61] for designing a robot developmental memory architecture. In principle, two identical robots are supposed to be influenced and guided by the same common knowledge from humans to generate two different emotions for the same stimulus based on their different past experiences.

The proposed requirements lead to build a new robot emotion representation based on human-like emotion activation cues : social effects and personal knowledge. In addition, a human-like cognitive memory such as Epigenetic Robot Intelligent System [61], is necessary to be adopted and developed. Currently, there is no integration of memory and emotions aimed at building a robot emotion model and providing robots possibility to develop intellectually in general. Although one fundamental emotion activation cue, social referencing, has been studied in the research of [74] to be integrated with a cognitive memory architecture to direct robot emotions and shape emotional response in a human-like way, robots are disable to link their formed object affective appraisals and generate emotions to a similar object based on past experiences.

For evaluating our robot emotion representation, a numbers of experiments are implemented by using two distinct robots which are provided by the same common prior knowledge through human guides. Visual images which contains several objects of a same type such as cat are given in a certain order to robots for each experiment. Besides, humans help to revise the object label, localize object positions in images and provide guide emotions. These robots then form object affective appraisals for each given image such as visual features to create a link for further usage. Robots are able to find the similarity between seen objects to the new object to gain experienced emotions and other engaged contextual information through retrieval phase. Both of guide emotions and experienced emotions influence the way robots generate personal emotions for the new object. After updating robot personal emotions, generated emotions are attached to the extracted information of the visual stimulus to be consolidated in robot long-term memory as an experience of the robot for readily accessing in the future. Besides, all robots are able to display personal emotions on two dimensional emotion framework [28] based on values of valence(negative-positive) and arousal(low-high) continuously through time. Furthermore, robots can display relevant gained experiences to a particular stimulus when successfully recalling those experiences from long-term memory due to human request. For instance, humans can ask robots to show their past related experiences from an image of a cat whose dominant color is yellow. Robots are able to show their personal emotions and other factors which influence robot emotion generation and knowledge acquiring processes. Several references from psychology experiments are used to evaluate our model in exploring the role of human-inspired emotion activation cues in robot knowledge acquirement and emotion generation.

1.4 Thesis Outline

Chapter 2 shows the literature review on this field of research and related works in modeling artificial emotions for robots. In addition, the limitation of the previous works are also addressed to propose our motivation on representing robot emotions.

Chapter 3 describes multiple emotion activation cues and their influences to human emotion generation and experience formation processes are described specifically. Moreover, this chapter also gives the basic idea of visual-based emotion generation in the development of human infants.

Chapter 4 aims at presenting our proposed model on representing robot emotions to deal with identified limitation of previous works. The design on robot long-term memory architecture is proposed based on the adoption of a developmental memory architecture to allow robots to remember experienced emotions. Furthermore, the modeling of social effects and the role of past experiences in emotion generation process from humans to robots are investigated.

Chapter 5 presents our experimental scenario and procedure. Based on the designed scenario, two main experiments are implemented to evaluate our propose model.

Chapter 6 draws the conclusion of this research contribution and future directions in robotic field.

Chapter 2

Literature Review

Emotions impacts to humans on the way of generating responses when being affected by the environment through a lot of processes to enhance memory and drive a cognitive mechanism for selecting behaviors. In addition, the personal emotion of an individual is influenced by not only their past experiences through memory retrieval processes but also people around. In attempt of creating artificial emotions for robots with the purpose of facilitating a believable interaction between humans and robots, a lot of researches were proposed and implemented different approaches. In order to study about emotions and how to model them for robots, we firstly need to define what are emotions and how to classify them.

2.1 Emotion Definition

Emotions are a complex function of humans which requires a lot of cognitive processes and physical change such as blood pressure, body temperature and heart beat during human development [60] [15]. Emotions can be defined based on positive or negative experience that is associated with a particular pattern of physiological activity. Emotions play a key role to create belief in a community based on emotional expressions and generated social emotions such as emphatic. Moreover, emotions help to reflex internal states, how a subject is affected by stimulus from the environment or others, and the way the subject adapts to the stimulus through emotional expressions, gestures and behavior; thereby improving survival and social skills of humans.

2.2 Emotion Classification

Due to the importance of emotions in human development, several psychology theories have been proposed to study and classify human emotions. In attempt to classify and represent emotion in various areas such as psychotherapy, psychology and cognitive science, researchers proposed to present emotions by two main approaches: emotions as discrete categories and emotions in multiple dimensions.

2.2.1 Discrete Categories of Emotions

The presentation of emotions based on discrete categories arises from theories aim at that different emotions are generated by different separated neural systems. In 1980s, Ekman defined basic categories of emotions which is limited at the small numbers of emotions: anger, disgust, fear, happiness, sadness, and surprise [21]. In the same approach, Plutchik presented another model which aimed at the extension of complex emotions based on the definition of eight

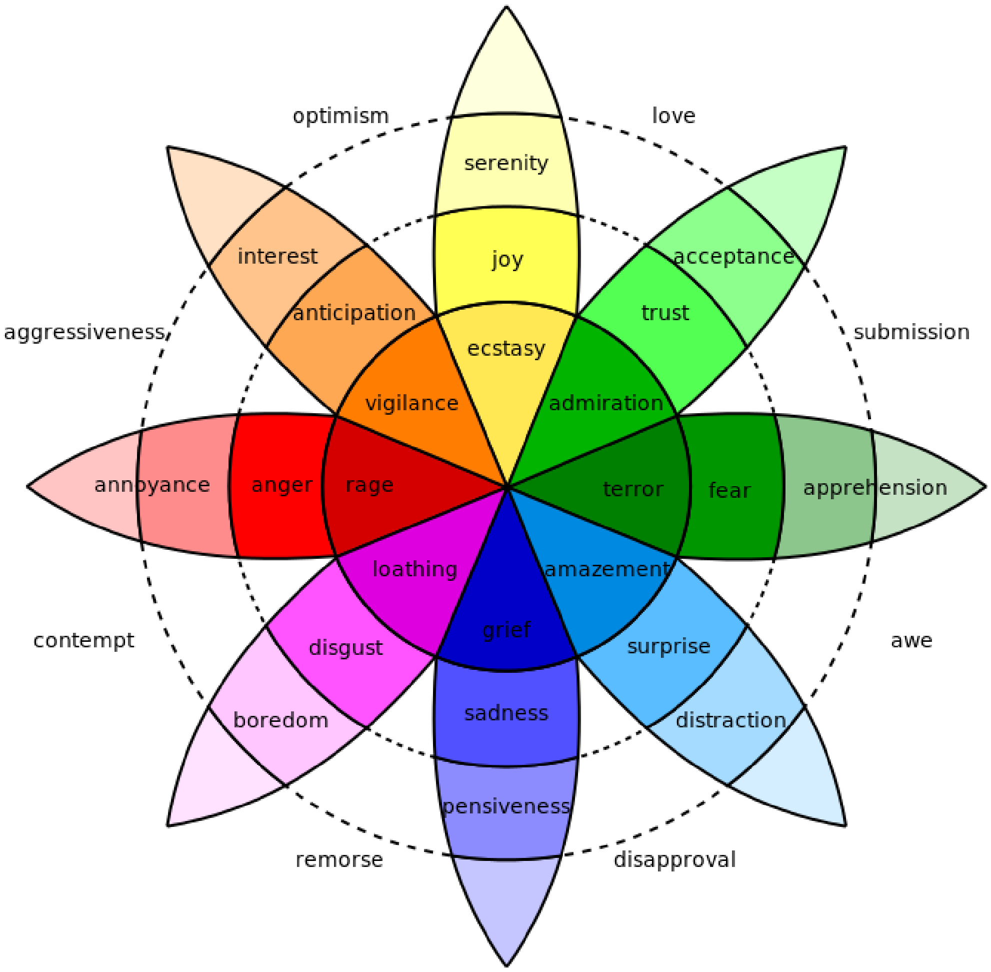


Figure 2.1: Putchik's wheel of emotions[1] which classifies the intensity level of an emotion in a class of emotions by the intensity level of the color of each emotion. White colored emotions are complex emotions created by combination of two emotions in eight primary emotions as shown in Table 2.1.

Table 2.1: Combination of two emotions in Putchik's model.

Combination of two primary emotions	**Result**	**Opposition**
Joy and Trust	Love	Remorse
Trust and Fear	Submission	Contempt
Fear and Surprise	Awe	Aggessiveness
Surprise and Sadness	Disapproval	Optimism
Sadness and Disgust	Remose	Love
Disgust and Anger	Contempt	Submission
Anger and Anticipation	Aggessiveness	Awe
Ancitipation and Joy	Optimism	Disapproval

primary emotions [26]. Primary emotions are grouped by four main couples of emotions: joy with sadness, anger with fear, trust with disgust, and surprise with anticipation. Different levels of intensity of a primary emotion can lead to the creation of other emotions. For example, joy can create ecstasy when the intensity level increases or generating serenity when the intensity level decreases otherwise. Furthermore, complex emotions can be formed by not only the

Table 2.2: A structural tree of emotions.

Primary Emotion	**Secondary Emotion**	**Tertiary Emotion**
Love	Affection	Adoration, Affection, Love, Fondness, Linking, Attraction, Caring, Tenderness, Comparison, Sentimentality
	Lust	Arousal, Desire, Lust, Passion, Infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, Bliss, Cheerfulness, Gaiety, Glee, Jolliness, Joviality, Joy, Delight, Enjoyment, Gladness, Happiness, Jubilation, Elation, Satisfaction, Ecstasy, Euphoria
	Zest	Enthusiasm, Zeal, Zest, Excitement, Thrill, Exhilaration
	Contentment	Contentment, Pleasure
	Pride	Pride, Triumph
	Optimism	Eagerness, Hope, Optimism
	Enthrallment	Enthrallment, Rapture
	Relief	Relief
Surprise	Surprise	Amazement, Surprise, Astonishment
Anger	Irritation	Aggravation, Irritation, Agitation, Annoyance, Grouchiness, Grumpiness
	Exasperation	Exasperation, Frustration
	Rage	Anger, Rage, Outrage, Fury, Wrath, Hostility, Ferocity, Bitterness, Hate, Loathing, Scorn, Spite, Vengefulness, Dislike, Resentment
	Disgust	Disgust, Revulsion, Contempt
	Envy	Envy, Jealousy
	Torment	Torment
Sadness	Suffering	Agony, Suffering, Hurt, Anguish
	Sadness	Depression, Despair, Hopeless, Gloom, Glumness, Sadness, Unhappiness, Grief, Sorrow, Woe, Misery, Melancholy
	Disappointment	Display, Disappointment, Displeasure
	Shame	Guilt, Shame, Regret, Remorse
	Neglect	Alienation, Isolation, Neglect, Loneliness, Rejection, Homesickness, Defeat, Dejection, Insecurity, Embarrassment, Humiliation, Insult
	Sympathy	Pity,Sympathy
Fear	Horror	Alarm, Shock, Fear, Fright, Horror, Terror, Panic, Hysteria, Mortification
	Nervousness	Anxiety, Nervousness, Tenseness, Uneasiness, Apprehension, Worry, Distress, Dread

combination of primary emotions with different levels of intensity but also the opposition of another emotion. Love which is the result of joy and trust, opposites to disapproval which can be obtained by the combination of sadness and surprise. The model and examples of combining one emotion with another emotion are shown in Figure 2.1 and Table 2.1 respectively. Recently, A tree structure of presenting basic emotions was proposed based on six primary emotions: love, joy, surprise, anger, sadness and fear for identifying more than 100 discrete emotions including a large number of secondary emotions and tertiary emotions with multiple different levels of intensity [56] as shown in Table 2.2

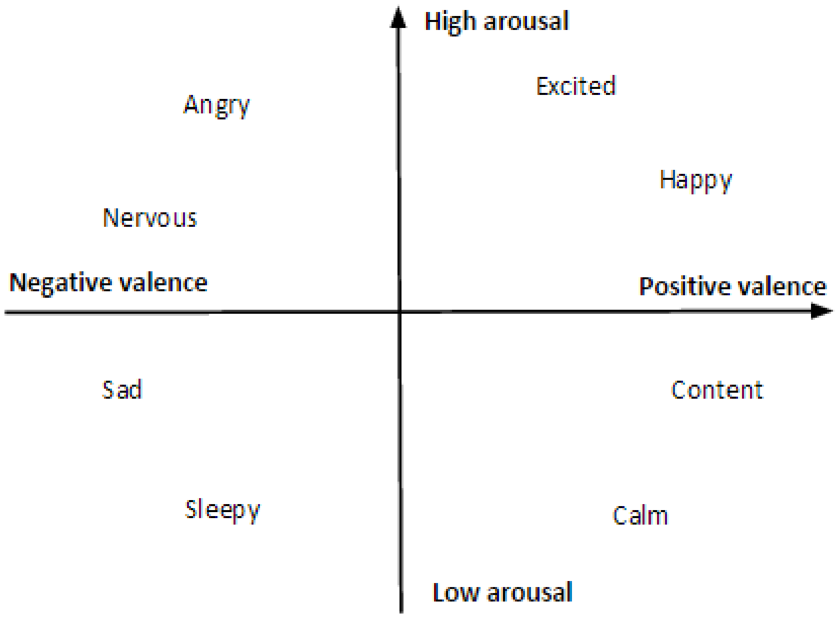


Figure 2.2: Emotions are represented in the affective valence-arousal coordinate.

2.2.2 Emotion Representation in Multiple Dimensions

Other researchers follows theories which are contrasted to proposed theories for discrete categories of emotions to shows that all affective states are related to a single interconnected neurophysiological system. They claimed that limited categories of emotions cannot discriminate human sufficient levels and proposed to represent emotions based on multiple dimensions. A long the line, 2-dimensional emotional framework [28] was proposed to present emotions based on two values of valence and arousal. While valence presents how positive or negative an emotion is, arousal shows the intensity level of the emotion. For example, happy and excited are positive emotions determined by different intensity levels as shown in Figure 2.2. Furthermore, human facial expression can be presented based on valence and arousal [66] which supports significantly in recognizing and reproducing human emotions and designing artificial emotional facial expressions.

However, this model were limited at discriminating the different of fear and anger and presenting social emotions. To deal with identified limitation of previous model, Russell(2003) [65] proposed core affects to construct emotions psychologically based on cognitive theories. Russell defined core affects as neurophysiological states consciously accessible as the simplest raw feelings evident in moods and emotions. At a certain time, the conscious experience (the raw feeling) is presented as a single point on a map of core affects as shown in Figure 2.3. The activation/deactivation which stands for the mobilization and energy of an emotional state, ranges from sleep to excitement. The pleasure/displeasure dimension which is raged from agony to ecstasy, presents for the condition of the emotional state. Two dimensions are similar to valence and arousal; thus, using core affects in the proposed valence-arousal coordinate to present emotions can effectively distinguish the different between fear and anger. Besides, we also can represent social emotions based on named areas in core affects; for instance, upset/distresses presents for bad feeling. Shame is a bad feeling of an individual about himself which can be the core affect presented in upset area of an cognitive element for the interpretation about the individual.

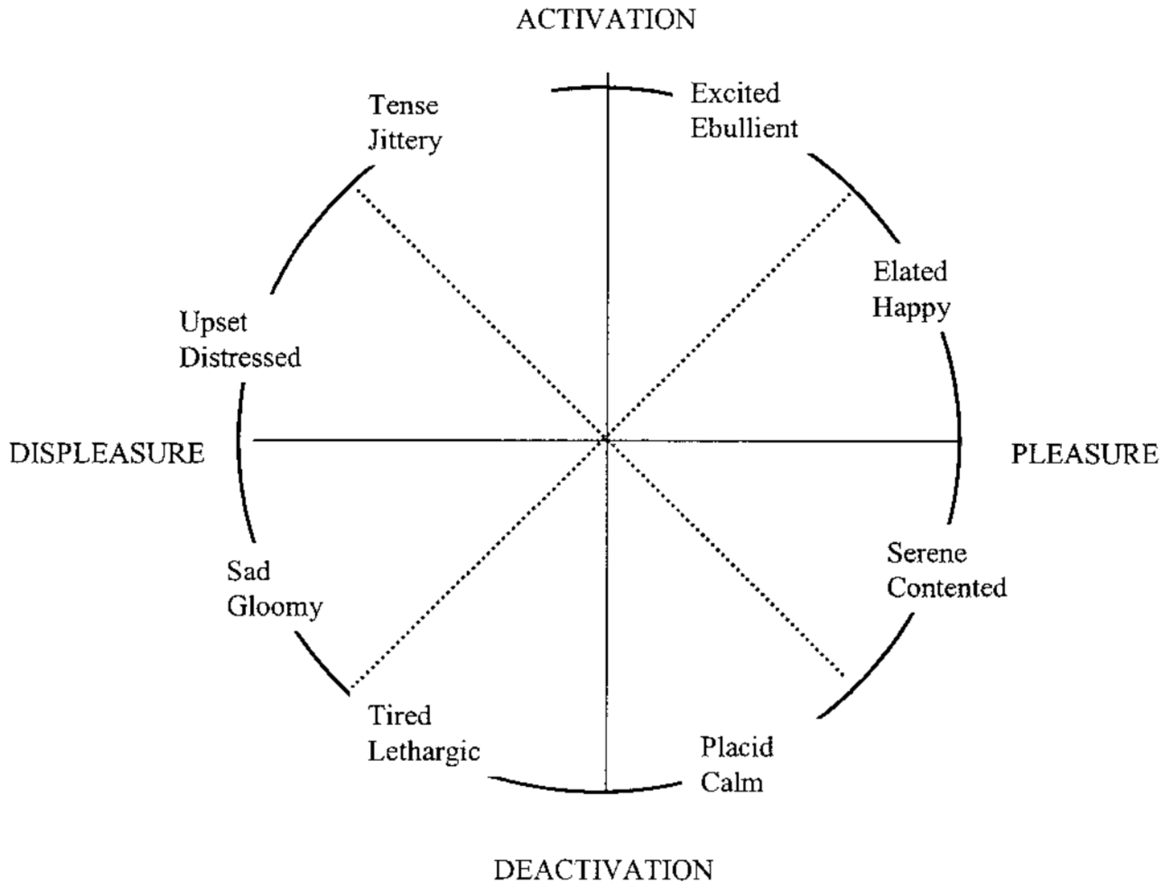


Figure 2.3: Core affects [65] help to improve the valence-arousal coordinate to discriminate fear from anger. Activation/deactivation and pleasure/displeasure dimensions are similar to arousal and valence dimensions of 2-dimension emotional framework respectively.

2.3 The Relationship of Human Emotions and Cognition

In humans, emotions are claimed to have strong influence in cognition, brain activities and cognitive elements such as memory, perception, attention and reasoning. This relationship may influence to the formation of personal interpretation about the environment and personality of an individual during human development process. The relationship also can help the individual to form detailed appraisals about stimuli or events for future usage in an ambiguous situation when being affected by an unknown stimulus. However, emotions can lead to disorders in memory and unreasonable reactions.

Emotions direct an individual to focus on certain features of a specific stimulus or situation, and evoke brain activities through changing physical components of the body such as skin, heart beat and blood pressure in order to react to that stimulus from the environment [54]. For instance, the individual can recognize a snake in the grass and generate fear emotion by some specific features of the snake such as body shape. Besides, the impact of emotions in attention do not maintain due to the reduced activity of prefrontal cortex followed by age [68]. Elders shows higher interference in non-emotional tasks and lower interference significantly during emotional task engagement in comparing with younger subjects.

In addition, emotions changed during memory retrieval, show the role in memory performance enhancement due to several factors such as mood, age, gender and emotional components of past experiences included valence and arousal of subjects. Remembering the event which has similar emotion with the current state of a subject is easier than other events with different emotions due to the effect of mood. The high intense of engaged emotions makes the event become strongly consolidated in human memory and easily to be accessed in the future while the age of experiences has stronger influence when those experiences are encoded with neutral low-intense emotions due to the effect of arousal component of emotion in memory enhancement. The subject may face to the same effect when remembering their past experience.

Furthermore, emotions assist human reasoning and shape behavior selection through decision making direction. Personal knowledge and emotions are generated based on formed affective appraisals and decision making processes during human development; thereby regulating reactions of an individual to the affective stimulus. For instance, an individual who is in a very emotional state such as lost a loved person, may not have reasonable thinking about another situation or stimulus such as studying or working. The individual then may score bad in studying or work ineffectively.

Cognition also helps to regulate emotions of an individual for analyzing external stimuli and internal past experiences. When being in a certain situation and affected by known stimuli, affective appraisals are required to recall past experiences including experienced emotions to change personal emotion of a subject; thus, the personal emotion of an individual is affected strongly [13]. When an individual is in a negative situation such as arguing with their parents, they can regulate their emotions to neither not think about the arguing nor express their personal emotions to friends.

2.4 Artificial Emotions in Robots

Due to the influence of emotions to cognitive processes and brain activities in humans, artificial emotions are modeled to robots for facilitating an effective interaction between humans and robots. Emotions can be used to help robots express emotional behaviors which can evoke interest of humans in sharing and social interacting through giving feedback to humans how robots are affected and adapt to stimuli from the environment during interactions with humans. Furthermore, due to the relationship between emotions and cognition in humans, artificial emotions can be used to drive attention, emotion generation and memory formation of robots to select behaviors for a certain task or situation. Implications of emotions in robots are presented in following sections.

2.4.1 Emotional Expressions with Social Robots

Emotions have been studied in designing and implementing social robots such as Paro [69] and Probo [67], for giving feedback about the internal state of a robot, how the robot is affected by and adapt to stimuli from humans or the environment. Social robots which can obtain information from provided sensors and cameras to generate emotions and express those emotions through gestures, behaviors and sounds. For example, Paro can get sensor information from a stimulus in the environment and react to the stimulus reasonably. When someone slaps the head of Paro strongly, Paro presents sad sound, lows down his head and closes his eyes to show his emotion is negative.

Those robots are provided basic functionality to generate and express emotions like humans; therefore, they can play a role as a friend or a pet for engaging with humans especially in children, elders and autism patients; thereby improving their social interactions effectively. Autism children are difficult to express their own emotions and do not want to engage in any interaction with others even their parents, and elders are lack of care from their family especially when they do not have their family anymore. The mental life and social interactions of those subjects are necessarily to be considered and improved. Social robots are designed to have good-looking; thus, humans show more interest when interacting with them [75]. When Paro shows negative emotion because of being slapped to an user, the user then evokes an emotion

(a) An animated Probo robot interacts with children[67].

(b) A Paro robot plays a role as a pet to interact with elders [69].

Figure 2.4: Illustrations of interactions between humans and social robots which can generate and express emotions.

such as empathy to interpret robot reactions and show more interest in future interactions. The interactions between social robots with children and elders are illustrated in Figure 2.4.

After a short-term of engagement, humans understand all characteristics and responses of those robots in all cases with different stimuli during interactions. The limited functionality disables social robots to generate adaptive emotions and expressions continuously due to time parameter. For multiple interactions with the same stimulus, a robot can generate only one emotion and respond due to that emotion, which can be observed after several interactions. For instance, Paro always expresses his negative feeling by lowing down his head and presenting sad sounds. Furthermore, multiple robots provided the same functions and interacted with different people can generate only one uniform emotion without considering about any past experiences or prior knowledge.

2.4.2 Integration of Cognitive Memory and Robot Emotions

Cognitive memory can provide robots with a capacity to acquire knowledge independently and continuously adapt to the environment, which can effectively helps to enhance a long-term human-robot interaction. Recently, this role of cognitive memory has been applied to model artificial emotions for robots, which helps robots to remember past experiences including engaged emotions for future recall in terms of the relationship of emotion and cognition in shaping emotion generation and directing behaviors [24]. An integration of a cognitive memory and emotions was proposed and implemented to a social robot, Nao, to increase the engagement between children and the robot during interactions [9]. Nao can recognize children emotional states to generate appropriate behaviors while a provided Episodic Memory helps to facilitate interactions between humans and robots for a long period. Therefore, the robot can acquire and maintain the interest of children after a long-time of engagement.

The role of emotions in driving behavior is applied for providing robots with a capacity to complete given tasks sufficiently. In a delivery task [44] as shown in Figure 2.5 , a robot needs to identify and remember the sender, a requested object and the message of the sender to detect the receiver and delivery location. The robot may fail several times and get nervous when trying to reach a receiver; thereby gaining emotional experiences for directing reactions to localize the receiver and deliver the object in the correct location. For another task, emotions

(a) The robot identifies the sender and the object.

(b) The robot gains experience from a wrong destination.

(c) The robot delivers the object to the receiver

Figure 2.5: A robot designed by the integration of episodic memory and emotions in delivery task. [44]

are implemented in an artificial intelligent agents to play a soccer game [5]. An agent can increase the percentage of accurate goals and number of scores during the match when keep kicking a ball; thereby generating emotions and gaining experiences to adjust behaviors.

Robots may acquire knowledge and generate emotions for reacting intelligently due to the awareness about situations. Robots gain knowledge and form affective appraisals through interactions with humans and self-exploration in the environment. The acquired experiences and knowledge through memory retrieval of episodic memory allow robots to evaluate and interpret unknown stimuli from an ambiguous event personally during exploration about the environment or interactions with humans, which may generate intelligent behaviors in response to the situation such as warning humans in a dangerous situation[20].

However, those robots can only simulate some specific human cognition to finish some given tasks, so they cannot do the same simulating cognition processes for other tasks effectively. A robot is designed for doing delivery tasks may not work sufficiently in cleaning tasks. In addition, the simulating of single component of human memory such as Episodic Memory does not help robots to conceptualize any given object, task, or situation in general in order to gain knowledge and reuse this information for future similar tasks or stimuli.

2.4.3 Robot Emotions with Social Context

During interactions with robots, humans can improve their social skills and increase their social interactions with others while robots can learn from humans via provided interpretation about the environment. This effect can be applied to the integration of cognitive memory and emotions for providing robots with a capacity of gaining knowledge and emotion generation with social context from humans.

Due to this effect, a robot were implemented based on the inspiration from human psychology about social referencing and biological theories, Leonardo [74] which can mimic human emotional facial expression as shown in Figure 2.6. This robot can detect human face and recognize human emotions represented in the valence-arousal coordinate[66]. The robot then can mimic human emotions and expression based on those values in respond to a specific stimulus. This stimulus is extracted some basic feature based on given with guided information during interactions between a human and Leonardo to form a novel personal affective appraisals of the robot for future usage. A toy is given to Leonardo with a fearful message from a human through human facial expressions. The robot detects that object through a share channel with the human, evaluates the message based on facial expressions to copy and reproduce the ex-

(a) The implemented real robot Leonardo based on human-inspired social referencing.

(b) Simulation version of Leonardo mimics some human emotional facial expressions.

Figure 2.6: Leonardo robot and some samples of the robot when mimicking facial expression to show robot personal emotions [74].

pression. Besides, some basic features of the toy are extracted by the robot through provided information from the human such as color and name of the toy. The robot can generate personal emotion for reacting emotionally based on the given message and encoded human facial expressions. All generated emotions, expressions and basic interpretation about the given object are encoded and maintained in robot memory. For the next interaction, when the human shows the stimulus to Leonardo with a neutral facial expression, Leonardo can recall experiences from previous interactions to generate the same emotions and reactions as the first time of seeing the stimulus.

In this integration of emotions and cognition with social context, the robot play only one role of a passive learner to accept human messages and emotional expressions directly through interactions. In other words, the robot is disable to give evaluation for similar experienced stimuli independently. It is necessary to consider the role of a partner for robots to learn from human guide and share emotions with humans during interactions.

2.5 Conclusion

We present the definition of emotions and how to classify them in this chapter. In addition, the relationship between emotions and cognition with social perspectives in humans is specified to claim that it is important to consider the role of cognitive memory and social effects in human emotion activation which are necessary to model artificial emotions for enhancing human-robot interactions. A lot of studies focused on those human-inspired emotion activation cues for modeling robot emotions to drive robot behaviors, which were limited at simulating some specific human cognition and expressing uniform emotional expressions due to a small number of defined emotions. Furthermore, the context of social effects was investigated based on social referencing and passive roles of robots during interactions with humans. However, this work disabled robots to generate emotions and acquire knowledge actively which can be done by providing robots with a capacity to learn intelligently from humans as a partner due to personal experiences.

Chapter 3

Human Emotion Activation Cues

Humans tend to behave emotionally to show their internal states to others and let others understand them deeply for making social relationships. During interactions with others, human emotions are related to not only the current stimulus, object or situation but also their past experiences which have several similarity to the current object. Besides, people around can influence to the personal emotion of an individual which can be considered as social effects to emotion generation of an individual. Hence, past experiences and social effects play important roles in activating emotions of an individual for a specific object, stimulus, or situation. In the following sections, we present the definition of those emotion activation cues in details with inspired psychology theories.

3.1 Social Effects in Emotion Activation

Social effects, an available resource of information with social context, are divided into two main parts which are social referencing and social sharing due to children developmental process. On the one hand, social referencing, which acts strongly in human infants, helps to teach an individual a basic knowledge about a specific object or a stimulus, and shape their emotions and behaviors in respond to the stimulus. On the other hand, social sharing helps the individual to generate emotions and drive emotional responses for a more detailed interpretation about the environment by using the acquired personal knowledge and experiences. Due to the interpretation of others, an individual can obtain personal knowledge, form personality, generate emotions and regulate reactions in respond to an ambiguous object, stimulus or event in order to adapt to the environment. In addition, not only basic emotions such as happy and sad but also social emotions such as guilty and embarrass are learned and shared by social effects. Following paragraphs show more details about each factor of social effects.

3.1.1 Social Referencing

Social referencing, which is the receipt and the use of other human's interpretation to form an individual personal interpretation about a situation and influence to the reactions in respond to the situation [22], acts at every stage of human development process especially in infants who is less then 18-month old. The preconditions for using and being influenced by social referencing are described in details by following parts.

Preconditions

In order to learn based on social referencing, humans are necessary to satisfy several preconditions about cognitive process and functionality which can be achieved in determined milestones

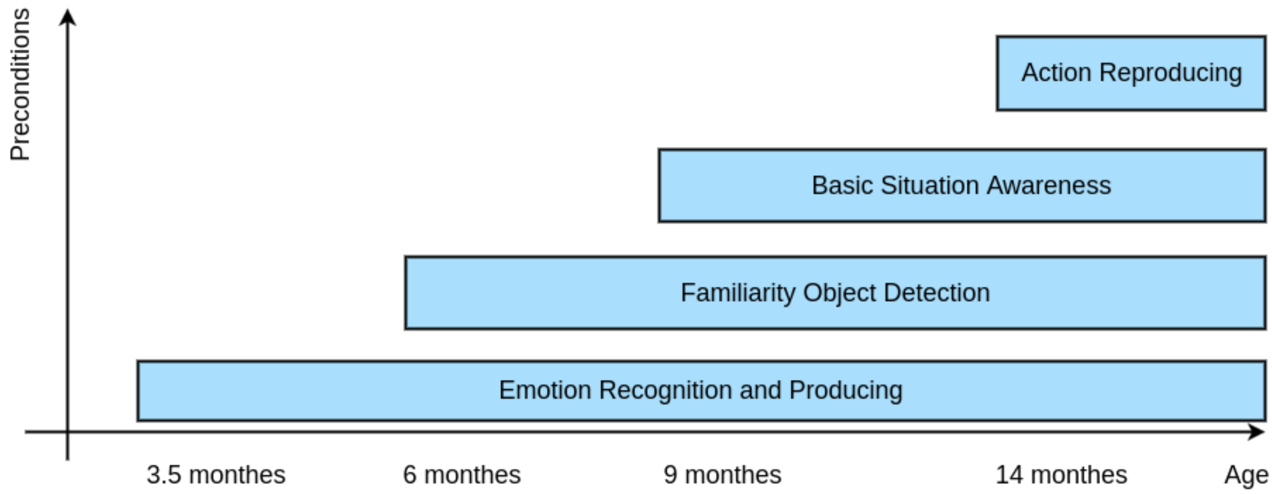


Figure 3.1: Functionality formation during human infant development to use and be influenced by social referencing.

in the development. Important milestones of infant development in context of social referencing are shown in Figure 3.1.

At age of 3.5-month old, infants start to have basic functions for recognizing binary emotions and emotional expressions of others such as sad and happy; thereby learning to mimic those emotions and the expressions due to those emotions.

At the next stage when infants reach to 6 months old, they are able to recognize several emotional expressions and respond differently to sad, happy, angry and fearful faces [77]. In addition, they can make some judgments about familiarity of those expressions to give different responses such as spending more time to look at their mother's face than others' faces when looking for a reference or a message in an ambiguous situation. This ability enable infants to mimic their parents emotional expressions in order to respond to stimuli in the future.

By 9-month old, infants are sensitive to evaluate the consequence of the reaction of a subject to a stimulus through looking at the subject's face [58]. It can lead to form awareness about the current situations through understanding the meaning of their parents' messages in order to generate emotions and reactions.

When being at age of 14-month old, infants are able to reproduce the contextual information of a sequence of two or three novel actions to an object they have seen and made appraisals through referencing to their parents [8]. This ability enable infants to make self -exploration about the environment without the presence of their parents to form personal knowledge and personality. For example, infants watched their parents turn on the light inside a woolen ball by putting their heads on the balls for several times, after that, those infants tried to do the same thing to see the light on. Therefore, infants can form their own affective appraisals for the stimulus and engaged social context including emotions when being affected by similar stimuli in the future.

The Influence of Social Referencing in Infants

Due to social referencing, infants rapidly accept the guide from their parents as a passive learner in acquiring knowledge about a certain object such as a toy, an optical illusions or a stranger. Infants generate emotions and behaviors in response to the stimulus by a mimicking mechanism to regulate their own emotions and behaviors as similar as the encoded emotions and expressions from their parents, which is presented in Figure 3.2. For the same toy and object appraisals, infants may have similar emotions and reactions for several times of interactions due to either the provided guide from parents or acquired knowledge [23].

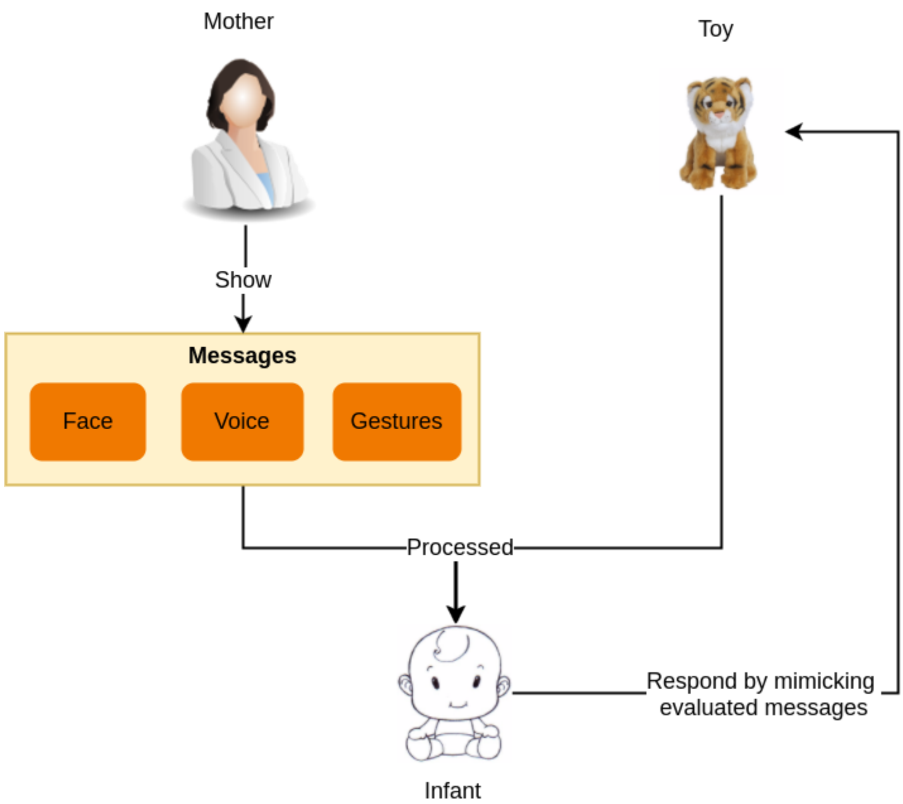


Figure 3.2: Social referencing in infants for generating emotions and expressing emotional responses by mimicking parent guide messages.

Infants use social referencing as a root of information resource to develop very basic personal knowledge; thereby forming object appraisals and generating emotions in respond to a toy due to copying emotions and expressions of their mothers and consolidating an experiences stored in their memory [22] [34]. At the first interaction with any stimulus, infants need to get guide from their parents to get interpretation about a stimulus and form an experience including engaged emotions to be consolidated in their memory. For next interactions with the stimulus, infants may use their past experiences to reproduce past reactions the toy without getting any guide from their parents, which can be observed in 14 month-old infants.

The way infants interpret parents guide is different due to the age of infants. The developed cognitive functions, social skills and the age of an individual infant lead to form different levels of detail of gained knowledge through interpreting the guide and affecting stimuli which can regulate emotions. Older infants can interpret messages from parents through looking at only the face of them while younger children look at their parents with no preference of faces or gestures to get parents guide [76]. In addition, younger infants spend more time than older infants when looking at their parents to get more guide for understanding one kind of emotional messages.

Based on the influence of social referencing, parents and caregivers can direct infants attention to a particular object or stimulus and shape their emotions in respond to the stimulus. Infants often look at their mother emotional expressions to decide for approaching to a new object, toy and person when getting a joyful message or staying far away from them by a fearful message otherwise. Smiles and soft voices of their mothers inspire them to get in tough with a stimulus while fearful faces and speeds generate a warning message to alert them stay far away from the stimulus [6]. Infants are able to show their emotions toward an object by emotional expression based on the guide from their parents such as smiling to their parents or crying in respond to strangers otherwise.

3.1.2 Social Sharing

Knowledge Acquirement

Social sharing takes a stronger place than other aspects of social effects at older stage of human development. When infants reach to 18-month old, they start to and negotiate with their parents guide on the way of gaining knowledge, forming appraisals and generating emotions [22]. Older children usually show more personality and independence when getting the guide from their parents than younger children who are mode involved to their parents. A child may ignore parent guide to move toward a toy when seeing the toy and have a thought of exploration expectation while a 6-month old infant stops moving toward the toy when getting a fearful message from their parents.

In addition, social sharing is an important learning method which helps an individual use the personal knowledge and experience to combine with the guide from others to gain a more detailed interpretation about the environment. When being in an ambiguous situation or affected by a new stimulus, the individual can use the past experiences and formed affective appraisals to find the link between the new stimulus and known experiences to evaluate the situation [27]. Moreover, the resource of information can be obtained from not only parents but also other families members and familiar stranger such as neighbors [35]. When in a situation which 2-year old children cannot have the presence of parents, they can get messages from strangers to understand the environment for making reactions for a certain object .

Social Emotions and Social Interactions

Furthermore, social emotions such as guilty, proud and sympathy can be generated by the influence of social sharing when an individual uses the personal knowledge to understand situations of another. The generated emotions, secondary emotions are distinctive from the primary emotions. On the one hand, primary emotions refers to a subject own emotion when being affected by the environment such as feeling fear when seeing a wild big animal or being sad when loosing a loved person [17]. On the other hand, secondary emotions are effected and transferred from other people during interactions with them. Shame emotion can be generated when a secret of a subject is well-known by other accidentally while guilty emotion might be created when an individual recognizes his bad behaviors to others. An experiment was implemented by [38] to give the evidence that preschoolers can generate social emotions. A story was given to preschool children about two characters who are Marry and her friend, and those children were asked to explain emotions of both characters in different situations. Marry got a bad experience about a dog who was chasing her dog in the past, and at the present, she sees the dog again she feels sad whether the dog waives his tail to her or destroys her flowers; however, her friend feels sad when being near Marry and otherwise feeling happy when being with others. Preschool children are able to explain the emotion of others when being affected by a certain stimulus through setting themselves as a subject to experience, sympathize, understand the character's emotional state.

High arousal emotion can elect higher chance of sharing among humans while valence component of the emotion can changes the climate of humans when sharing. In the experiment of [11], participant were asked to watch an emotional videos and answer several questions of willing to share the experience of watching video with their parents or friends. In case of watching high arousal regardless to how negative or positive emotion the video brought, participate were willing to share with their presents. Positive emotions help to encourage humans around to share their positive experiences while negative emotions evoke their negative experiences, thereby regulating their personal emotions become more negative.

3.2 Personal Experiences in Generating Emotions

Emotions have strong influence on the way an experience is processed and transformed through memory encoding, maintained through consolidation phase, and relived through memory retrieval [51]. An individual can experience emotions, and then forming affective appraisals related to the stimulus including engaged emotion to be consolidated inside long-term memory. This emotion can be relived to let the subject face to the same effect from the original experienced emotion which can effect to the personal emotion at the current time.

3.2.1 Experiences of Emotion Formation

A lot of studies about emotions with multiple theories of emotions in various areas were presented including appraisal theory of Lazarus(1991) [43]. This theory aims at that the generation of emotions in humans in respond to a certain stimulus is following by thinking about internal and external factors involved. In other words, when being affected by a certain stimulus in a situation, an individual's physical components change due to the stimulus; for instance, heart beats and blood pressure of an individual increases when seeing a big spider, and body temperature of the individual decreases when seeing snow. The obtained personal knowledge are recalled due to some external factors. An image of happy people while building snowmen can evoke all past experiences of an individual about snowmen and snow. Combining external factors with internal appraisals, emotions of the individual are generated. If the recalled past experience is involved with happy emotional state, the subject can relive the experience and regulate to change his current emotional state positively.

Experiences about emotion are enriched with content details during the interaction of an individual with the environment involving to core affects, the intense of those affects and the change of internal and external factors through brain activities [7]. The core affect information which includes the pleasure of emotions of an individual to the interpretation of external factors, impacts to the internal cognitive processes of an individual. The individual then can interpret the danger of an situation, a object or a stimulus for further reaction regulation such as approaching or distancing. Based on the awareness about those stimuli, the individual generates perceptions and conceptual knowledge about the environment in order to produce personal emotion state for adaptation automatically to caused stimuli. In additions, the intensity level of emotions, which effects to the way of generating and experiencing emotions, allows subjects to understand better about the environment and sharpen knowledge. Combining both factors of core affects and the intense of emotions, a subject can experience emotions with different level in the same pleasure such as sleepy, angry, sad and bored which belong to negative emotions.

Lacking of accessible emotion concepts, the meaning of affective experiences and perceptions are ambiguous regarding to different prior knowledge and interpretation of humans about a certain stimulus. An experiment is implemented to observe brain activities while presenting emotional words such as anger and happy and general words like pleasure to subjects [14]. The activities of regions related to semantic processing are increasing significantly when subjects are presented by emotional words while the same effect did not appear in respond to general words. This result shows that the presentation of different detailed meaning of different intense of emotional words lead to different representation in perceptions of subjects due to their different prior knowledge. Therefore, generated emotions to the same general word are different for distinctive subjects due to their past experiences and personal knowledge.

3.2.2 Emotions in Memory Enhancement

Valence which refers to how negative or positive the emotion is, can act individually for facilitating memory performance. Both negative and positive valence allows the engaged experiences are consolidated and maintained inside memory for longer time than other experiences integrated with neutral emotion. Experiments with word recognition ability of subjects were implemented while controlling a large variety of lexical and semantic factors in order to evaluate the influence of valence in memory enhancement [3]. Subjects were able to remember positive and negative words regardless to arousal factor. Besides, for the greatest influence of emotions in memory retrieval phase, valence can influence to memory performance differently for positive emotions and negative emotions, in which positive emotions play a larger role during Episodic Memory search while negative emotions do the same during elaboration in Episodic Memory [25]. Negative emotions lead the subject to focus on some specific detail of the stimulus, then encoding and consolidating an experience based on focused information. Positive emotions do not show strongly effect on directing attention of the subject, then all contextual information of the stimulus is encoded and consolidated effectively in LTM. For retrieval phase of this experience, positive emotions act strongly to recall all related information of the original experience while negative emotions do not show a similar impact. Therefore, positive valence helps the subject to remember detail contextual information of the event more than negative valence.

The intensity of the emotion, arousal, regardless to how negative or positive the emotion is, shows influence on memory performance enhancement. Arousal component of emotions helps to increase interaction between amygdala and hippocampus in not only humans but also animals; thereby enhancing encoding and consolidating memory processes of arousing information [19]. Humans can recall past experiences more effective in high arousal condition than in non-arousing condition. Besides, arousal acts strongly in directing the attention of subjects to encode central detailed interpretation in memory encoding and consolidation processes, which is mostly relevant to the emotional content of the focused features of a stimulus or situations [37]. For instance, a person focuses more to a knife when seeing this knife towards himself rather than the person who is holding the knife. The directed information is better encoded and consolidated in long-term memory for future retrieval than other contextual information. In high arousing condition, the person can easily recall the experience of the knife without any detail of the knife holder.

Not only one individual component of emotions but also the interaction between them can influence to enahnce memory performance. The influence of valence depends on high arousal when an experience is recalled [41]. Neuroimaging studies also confirmed that the engagement between amygdala and orbitofrontal is strengthen for positive and negative emotional experiences in high arousal condition but brain activities for positive emotion factor are different from activities for negative emotion factor [51]. In high arousal condition, experiences with negative emotions are more likely consolidated and recalled vividly than others with positive emotions. In contrast, in low arousal condition, the influence of negative and positive valence are not maintained, which can be considered to the influence of single component of emotion, valence. Similarly for negative events, high arousal enhances memory better than neural arousal. Therefore, both components of emotions, valence and arousal should be considered equally when studying about the influence of emotions in memory enhancement.

3.2.3 Relive Past Experiences

Knowledge is acquired during the interactions an individual with the environment or others. The knowledge may or may not engage with an emotion about a certain stimulus or situation. Due to the influence of emotions on memory performance enhancement, the obtained knowledge can be consolidated vividly for reliving the original experience. Besides, when a certain

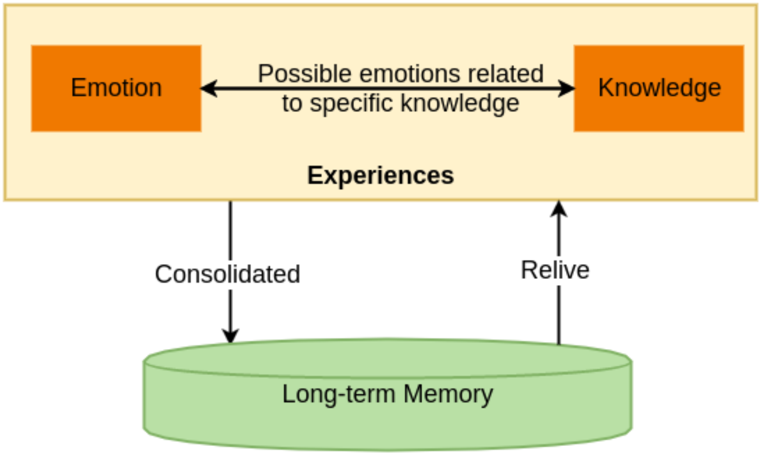


Figure 3.3: Reliving past experiences consolidated in long-term memory.

experience is recalled with the engaged emotion, the subject may face to the same effect as the original experience did. The reliving past experiences after memory consolidation is presented in Figure 3.3.

Through memory retrieval, humans use past experiences as a personal resource of information in order to continuously learn from and adapt to the environment. This resource is formed and updated through memory phases depending on new experiences and changes in the environment [19]. Therefore, affective appraisal which are generated for every object through interactions with others, can be used to make evaluation about the unknown objects to find the similarity between seen objects with the new object to generate emotions and regulate behaviors to gain survival skills. For instance, a child get pain when hitting a wall. He can make his evaluation on the wall such as position, color, texture and shape based on his past experiences about those knowledge to form the appraisal for the wall and encode acquired interpretation to consolidate a new experience including his personal emotion, sensory and body manifestation in his memory. The next time he sees the wall, he recalls his past experiences and relives experienced emotions to make the evaluation about the wall which may enable him to stay far way from the wall to avoid getting the same pain from the past.

The personal emotion of an individual is changed by the influence of reliving experienced emotions through retrieval process of memory. When recalling past experiences, humans recollect the original experience including all engaged sensory, contextual information, semantic and emotions [16] based on the relationship among long-term memory components, which leads to change subjects' current emotional state [4]. Preschool children may face to a negative effect when they relive experienced emotions of bad experiences through retrieval of memories [38]. College students recalled vividly autobiographical conceptual information and perception including emotions for high-arousing experiences regardless to the age of experiences [73], which may lead them to change their personal emotions.

3.3 The Interconnection of Emotion Activation Cues

Social effects helps humans to generate emotion and acquire knowledge with social context while past experiences provide humans with a capacity to have their own emotions which may different from others in respond to the same stimulus or situation from the environment. The external available resource of knowledge with social context which is provided by social effects, helps to update the personal experiences and form personality during the development of humans. Children at age of 3-year old are able to link their past experience to the current stimulus to generate emotions [39]. In integrating the evaluation and the guide from others such

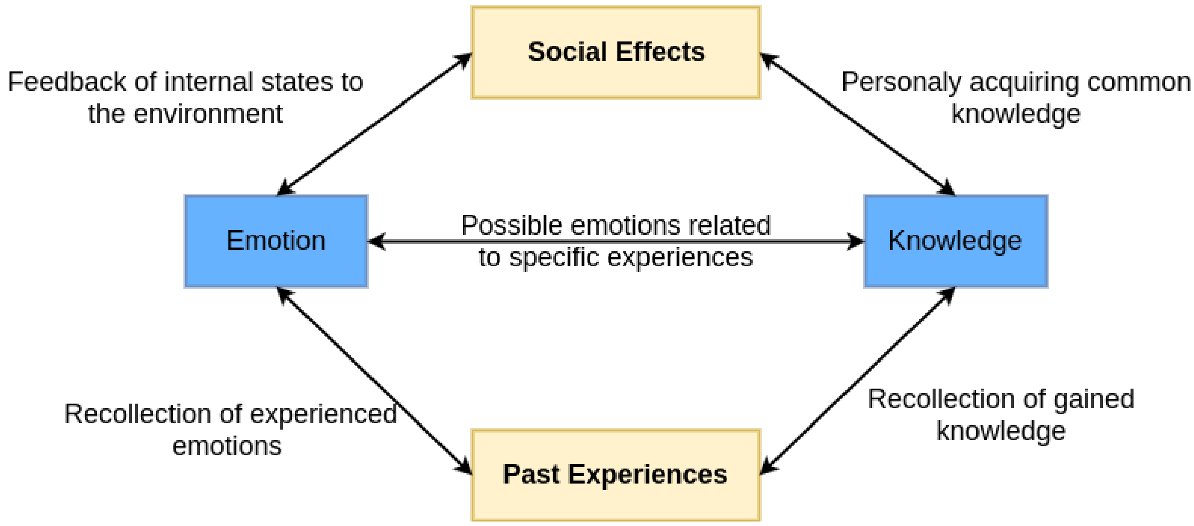


Figure 3.4: The interconnection between social effects and past experiences in generating emotions and acquiring personal knowledge of an individual.

as parents, the person can generate emotions and express them to adapt to the environment in respond to the stimulus. For instance, a toy of a child is broken by his mother's friend; and on the birthday of the child, his mother's friend brings a present for him but he feels sad when thinking about his broken toy. However, his mother tells him that the friend is really sorry so he brought a new toy for the boy, then the boy feels happy again.

Moreover, the acquired contextual knowledge of an individual is activated and strengthened while the engaged emotions are faded when being recalled to share with others [57]. For example, children usually talk to their parents or relative about their daily activities including negative experiences, thereby practicing to remember details of the experiences including the engaged emotions. Thus, those negative experiences are maintained vividly in children memory. Besides, their current emotions are changed while the experienced emotions are updated due to the new experience through interacting and sharing with others. In contrast, the emotional experiences are less likely faded when the subject does not share with others even though the contextual information is not vivid as the original experience.

Both factors are interconnected to each other through the development of an individual. Social effects provides an individual with an available accessing common knowledge and regulates personal emotion of the individual. The regulated emotion is engaged to a certain knowledge for forming experiences consolidated in long-term memory and relived through memory retrieval. The relived experience including related emotion may form personality of the individual in acquiring the common knowledge and giving feedback about the internal states of the individual to the environment when being affected by stimuli from the environment. The interconnection between social effects and past experiences in regulating the way an individual acquires knowledge and generates emotions are illustrated in Figure 3.4.

Furthermore, when sharing emotions and past experiences to others, an individual not only changes the personal emotion but also effects to change personal emotions of others. Humans tend to share experiences to others especially to people in their close relationship such as family members and friends. During interaction with other, the subject's personal emotion is effected by the recalled emotions and emotions from others. In additions, social effects allows shared humans to interpret shared interpretation and emotions based on their past experiences. Therefore, shared subjects can generate social emotions such as sympathy to get the same emotions by putting themselves as the subject to face to the shared experience [39], which leads to change their personal emotions.

3.4 Conclusion

This chapter presents fundamental human-inspired emotion activation cues during human-human interactions and the interconnection among them. Social effects help human to develop personal interpretation and generate emotions to for a certain stimulus in the environment based on the guide and knowledge of others. Two perspective of social effects, referencing and sharing, are described due to milestone of human development. Social referencing helps human infants to mimic emotions and acquire basic knowledge from parent guide while sharing aspects of social effects helps children and adult to generate a more complex detailed interpretation about the environment and generate personal emotions. The gained knowledge are consolidated in long-term memory as past experiences for future recalls to react to ambiguous situations and unknown stimuli.

Chapter 4

Robot Emotion Representation based on Multiple Activation Cues

Our robot emotion representation aims at integrating the influence of two human-inspired emotion activation cues, social effects and past experiences through memory retrieval process presented in the previous chapter. Robot personal emotion is updated continuously due to time parameter and the influence of modeled activation cues in respond to a certain stimulus from the environment. We begin this chapter by providing the overview of our robot emotion representation.

4.1 Approaches

Human emotions which are learned from and shared with others due to stages in the development, show personality of an individual in respond to stimuli affected during interactions with others and the environment. As presented in the previous chapter, we ague that the similar mechanism can be applied in robots to facilitate human-robot interactions through generating human-like emotions since humans tend to treat robots by the way they treat others.

The influence from others via social effects in humans can be modeled as guide emotions of humans for robots in order to interpret and respond to stimuli from the environment during human-robot interactions. In infants, they interpret their parent messages by evaluating facial expression, voice and gesture to recognize emotions. Thus, we can conclude that infants may have the same emotional state when they imitate their parent behaviors in order to respond to affective stimuli. For children and adult, their emotions can be affected by others through experience sharing due to social effects mention in Section 3.1.2. Due to the emotions from others, the personal emotion of an individual can be regulated for more acceptable possibility by others and the adaptation to the environment. Therefore, we model the role of robots to become an individual who is guided by humans to generate emotions which would be more favorably accepted by humans. Robot personal emotions are influenced by guide emotions due to the role of a passive learner and an active partner robots play based on social referencing and social sharing respectively. Thus, robots can generate emotions appropriately to adapt to stimuli from the environment during self-exploration or interactions with humans.

In addition, past experiences which are recalled during interactions of an individual with others, can be influenced to emotion generation and present personality of the individual. Human infants can form their own affective appraisals for a novel object and generate a specific emotion for that object due to the interpretation provided by their parents. When infants are affected by the same object which they experienced previously, they can recall their past experiences and respond emotionally as similar as the recalled knowledge including experienced emotion. For children, their knowledge is created by not only the acceptance of the interpretation pro-

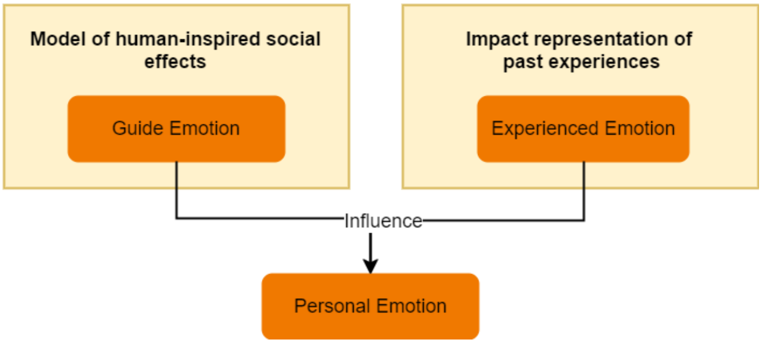


Figure 4.1: The model of human-inspired emotion activation cues to robot emotion generation.

vided by others but also their own personalities, past experiences and emotions. Thus, when reliving their knowledge, past experiences and experienced emotions, they can acquire different understanding and generate different emotions in respond to a certain stimulus. Due to the influence of past experiences to personal emotion of humans, we provide robots with a capacity of remembering past experiences through memory retrieval to relive experienced emotions which can influence to robot personal emotions.

Therefore, robot personal emotion generated for responding to a certain stimulus from the environment during interaction with humans can be influenced by guide emotions and experienced emotions as presented in Figure 4.1. Guide emotions provided by humans which is modeled based on human-inspired social effects, uses for directing robot emotion which would be more favorably accepted by humans. Experienced emotions which presents for the influence of robot past experiences through memory retrieval, can help to discriminate multiple robots with the same functionality.

4.2 Obtaining Guide Emotion

We propose to model human-inspired social effects through giving robots guide emotions which are generated by humans when being affected by the same stimulus given to robots. We model guide emotions for robots based on two main approaches: using a deep learning technique to enable robots to recognize emotions from given visual stimuli and getting human guide emotions directly through human-robot interactions. We firstly present the first approach which models guide emotions for robots through a convolutional neural network.

4.2.1 Guide Emotion Recognition based on Convolutional Neural Network

Motivation

Machine Learning techniques have been developed to enable computers to learn general rules for mapping inputs to desired outputs from a given desire data. Among those techniques, Convolutional Neural Network (CNN) were proposed for recognition, regression and classification tasks with multiple kinds of input data such as visual stimuli. CNN can help to automatically extract features from a raw image to learn and map those features to desired ground truth of outputs while there are a lot of other techniques, humans have to extract futures manually such as Support Vector Machine (SVM) when good features are not easy to obtain. Due to this benefit of CNN in visual recognition tasks, we apply this technique for enabling our robots recognize human guide emotions. Firstly, we present about the general concept of Artificial Neural Network to describe how CNN works and the way we apply CNN for solving our identified objective.

Artificial Neural Network

Artificial Neural Network (ANN) were developed based on neuron network in human brains [63] [79]. An ANN contains three main kinds of layers: input layer, hidden layer and output layer. In classical ANN architecture, there are always one input layer and one output layer while the number of hidden layer depends on the objective of specific cases. Each layer of ANN include at least one neuron unit. Each unit in a layer connects directly to every neuron unit in the next layer. The link between two units is modeled to present for an aspect of the complex connection of neurons in human brains. The illustration of an ANN structure is shown in Figure 4.2.

In general, before applying an ANN for doing any task, input data are processed to get the same domain of ranges. For example, when we extract features of a house such as width or height for the objective which is measuring how much money we can gain when we sell it. Width and height should be preprocessed to have the same measurement unit such as meter. After preprocessing processes, input data is put in the network. At each neuron unit for every layer except input layer, a function is used to activate the neuron based on the influence from all connected neuron units in a previous layer and the bias as presented in 4.3. The influence of neurons in the previous layer to a neuron in the current layer is calculated based on a weight parameter which is randomly initialized within a small range. The neuron unit after activating is used to integrate with a corresponding weight for feeding forward to neurons in next layers until all neuron units in output layer are reached. All neurons in the output layer are used to compare with ground truth values which are desired outputs to get the error. The goal of this network is minimizing the error between the output of the network and the desired output. In other words, a cost function is designed based on the comparison between output of the network with ground truth for obtaining the minimum error between those values. The error is moved backward to previous layers to update weight parameters, which is back propagation. The flow of feeding forward mechanism and back propagation is presented in Figure 4.4. In an ideal case with optimized algorithms, the cost function converges to a minimum point after a certain of training time. The point shows a threshold of the network for mapping inputs to desired outputs and the performance of this network.

In image processing, when trying to extract abstract features for visual tasks related to emotions manually, a particular idea were proposed, in which all pixels in a image are considered as a feature to be given into the input layer of ANN. The network structure of ANN becomes extremely big and it lead to increase the complexity and reduce speed performance of the network. For instance, we have a data consisted of 64x64 gray-scale images. The input layers of ANN contains 4096 neurons and each neuron is fully connected to all neurons in the next layers. When increasing the number of input neurons, we may need to increase the number of neuron in hidden layers and the number of hidden layers to extract good features. In other words, the complexity of the network increases with a large number of parameters which may lead to reduce the speed performance of the network. However, even considering such a complex network architecture based on classical ANN, extracting abstract features for visual tasks related to emotions are remaining challenge. Thus, it leads to the idea of proposing Convolutional Neural Network based on classical ANN to deal with such tasks for extracting automatically good features, and especially supporting for visual tasks regarding to emotions.

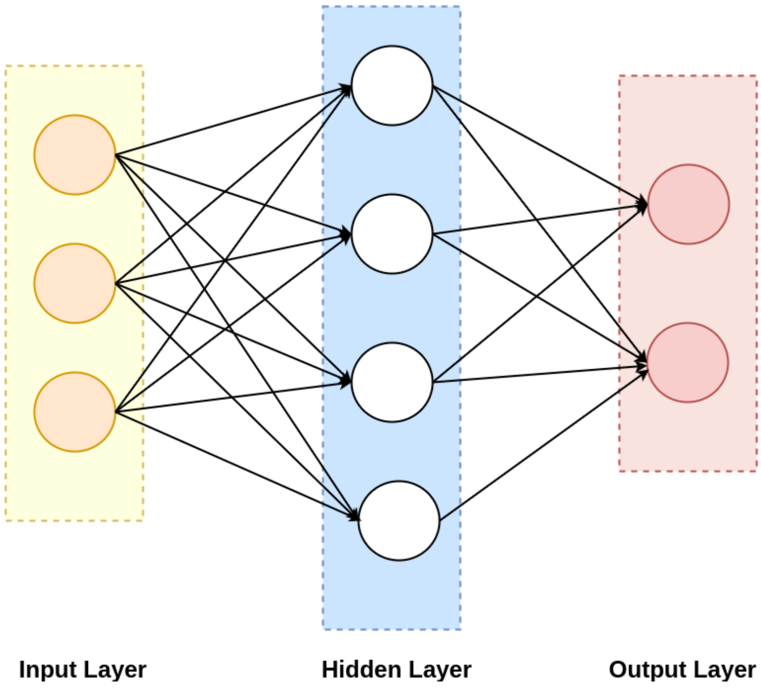


Figure 4.2: An example of structure of a fully-connected Artificial Neural Network. The number of hidden layers in an ANN and the number of neuron unit in each layer can be changed due to specific cases.

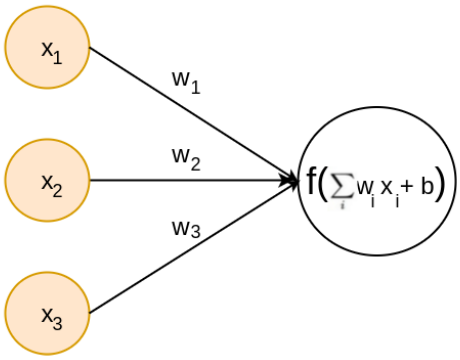


Figure 4.3: A representation of the input processing in one neuron based on an activation function for all influence from neurons in the previous layer and a bias parameter.

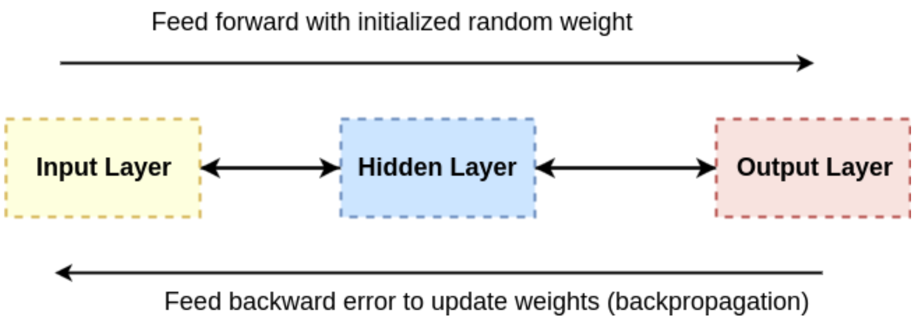


Figure 4.4: Weights between layers are initialized randomly in a defined scale to feed forward the information about features from the input layer to the output layer. After getting neurons in the output layer, activated values are compared with desired outputs to get the error between them, and this error is sent back ward from the output layer to the input layer for updating the weight and minimizing the error.

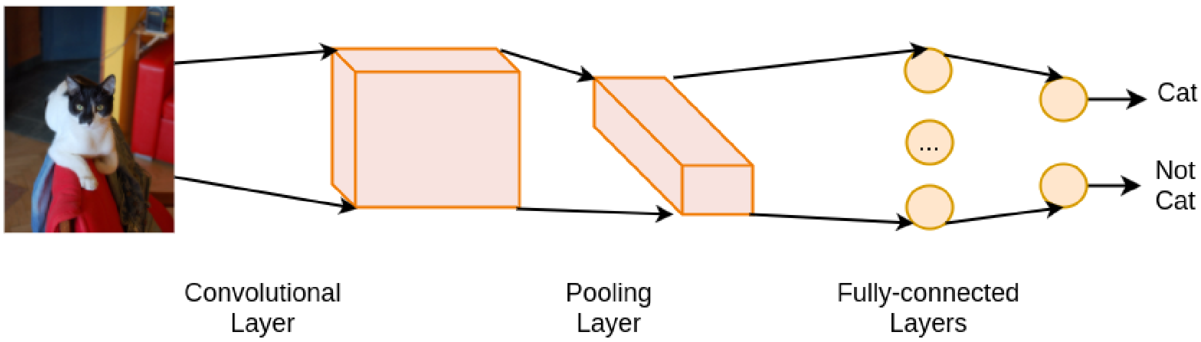


Figure 4.5: An example of a Convolutional Neural Network structure for cat recognition task included one input layer, one convolutional layer, one pooling layer and two fully-connected layers.

Convolutional Neural Network

In attempt of autonomously extracting good features such as abstract features from raw images, CNN were proposed and analyzed based on the idea of ANN [36] [45]. Instead of extracting features manually, extra layers are added into the classical ANN to form a CNN architecture. Those layers include convolutional layers and down-sampling layers (pooling layers) to extract features automatically. Besides, the architecture of CNN can help to reduce the number of parameters in the structure of a network when the input layer contains raw images. The architecture of a CNN is illustrated in Figure 4.5 for object recognition task.

Convolutional Layers are the core of the CNN which contains the most heavy computations and helps to extract features from input layer based on learn-able filter maps. Each filter map has a small size whose height size and width size equal to each other. For example, a filter map can have a size of 7x7x3 for color images have width and height equal to 7 pixels, and 7 pixels respectively in three channels: red, green and blue. Filter maps slide over the input image to compute the dot product of the filter map and corresponding region which has the same size with the filter map of the input image. The dot product is calculated as similar as matrix multiplications. Since we slide the filter over the width and the height of the input image, we produce a 2-dimension output which is the result of the dot product. This product is sent forward until the final layer of the network to obtain outputs for comparing with the desired values. This network structure also can do back propagation to feed backward the error obtained after comparing output of the network with ground truth to update parameters in filter maps. Therefore, filter maps can be updated to automatically to extract good features due to specific tasks. For example, a CNN can easily extracts edge feature based on a certain given input data due to learn-able parameters in filter maps as presented in Figure 4.6. As a result, the network can automatically learn from the filter map to extract features for mapping input data to expected output. The number of filter maps are flexible due to the objective of certain tasks for each convolutional layer. Thus, we can extract more features from the original input image. However, the large number of filter maps increases the complexity of the network.

The convolutional layer has some main characteristics as following:

- Receive an input which has a size represented in three dimensions: width, height and deep denoted as W_1, H_1, and D_1 respectively.
- Contain four main parameters:
  - The number of filter maps (K).
  - The size of all filter maps (F).

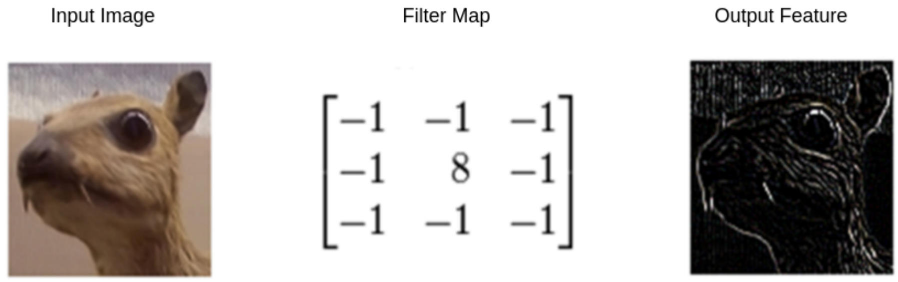


Figure 4.6: Extracting edge feature based on a learn-able filter map of a convolutional layer which slides over the input image.

  - The stride (S) which filter map moves over the input.
  - The amount of zero padding (P) for controlling the size of the output of this layer.

- Give an output which has a size of (W_2xH_2xD_2) after doing a dot product of filter maps and inputs.
  - $W_2 = \frac{W_1 - F + 2P}{S+1}$
  - $H_2 = \frac{H_1 - F + 2P}{S+1}$
  - $D_2 =$ K
- Use ReLU usually as activation function after getting the output: $f(x) = max(x, 0)$.

Pooling Layers are usually inserted after Convolutional Layers to progressively reduce the number of parameters and complexity of a CNN structure. These layers helps to control the over-fitting problem of the network. In other word, pooling layers can reduce the match between extracted features from convolution layer and output of the network for the train data. Therefore, it may enable the network to be more flexible to fit for other data in the future. Pooling layers contain several characteristics as follows:

- Receive an input which has a size represented in three dimensions: width, height and deep denoted as W_1, H_1, and D_1 respectively.
- Contain 2 main parameters:
  - The size of the spatial extend (F).
  - The stride (S).
- Give an output which have a size of (W_2xH_2xD_2) after doing a dot product of filter maps and inputs.
  - $W_2 = \frac{W_1 - F}{S+1}$
  - $H_2 = \frac{H_1 - F}{S+1}$
  - $D_2 = D_1$
- Use Max function or Average function as an activation function after getting the output.

The pooling layers use an activation based on Max function or Average function to reduce the size of an input. Max function takes the largest value in a identified region while Average function produce the average value of all values in the defined region. Figure 4.7 shows how Max function works in a Pooling layers which takes 4x4x1 input to produce 2x2x1 output based on the spatial extend F = 2 and stride S = 2.

Fully-connected Layers of CNN are similar to ANN which each neuron in the previous layer connects to every neurons of the next layers.

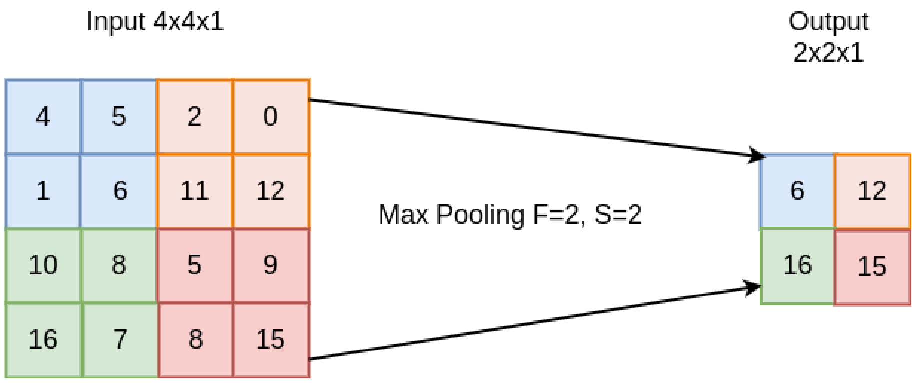


Figure 4.7: Max function acts in a pooling layer to reduce the size of the input based on a spatial extent F and a stride S on a defined region.

Implementation

Structure of The Designed Network

We design our own CNN to enable our robots to recognize emotions from natural images based on linear regression. Robots are able to generate emotional values(valence and arousal) which are represented in a continuous rage from input images. The extracted inputs are given in our CNN whose architecture is presented in Figure 4.8.

In the first convolutional layer, we have 32 filter maps while the second convolutional layer has 72 filter maps. The size of every filter maps in two convolutional layers is 5x5x3 similarly. Besides, we use LeakyReLU as our activation function for convolutional layers due to the its benefit of dealing with problem of dying ReLU which leads all weights equal to zero during training the network. LeakyReLU can control weights of convolutional layers through the parameter α given by Equation 4.1.

$$f(x) = \begin{cases} \alpha \cdot x \text{ if } x < 0 \\ x \text{ if } x >= 0 \end{cases} \tag{4.1}$$

Each convolutional layer is followed by a pooling layer to reduce the complexity of the network. Our pooling layers use stride S = 2, spatial extent F = 2 and Max function. In addition, we use two fully-connected layers after the last pooling layer to map extracted features to the desired network. The first fully-connected layer uses 90 neuron units and tanh as an activation function while the second layer use ReLU as activation function and contains only one neuron unit. The reason is because we want our output contain only emotional values which is either valence or arousal and the scale of valence and arousal included in our data set is from 1 to 9. Besides, our network is trained from scratch which initializes weights randomly in scale of (0,1). The cost function is used for this network to train our data set is built based on L2 normalization given by Equation 4.2.

$$lost = \sum_{i=0}^{n} |f_i - y_i| \tag{4.2}$$

Where:

- n is the number of data samples.
- f_i is the value of output neuron of the i^{th} sample of training set.
- y_i is the desired value of the i^{th} sample of training set.

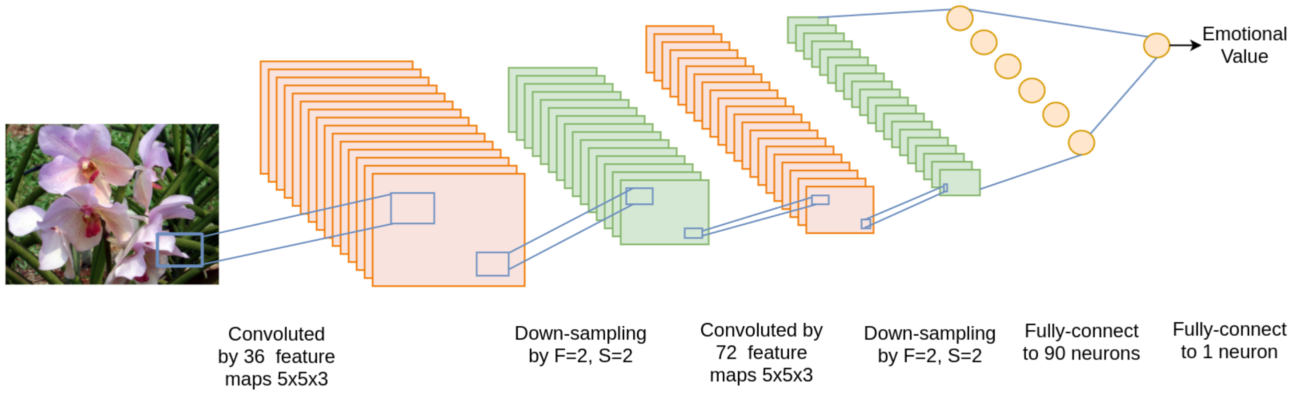


Figure 4.8: Our proposed CNN architecture to recognize emotional values from one stimulus based on Linear Regression.

Experimental Data

International Affective Picture System (IAPS) [42] were developed based on examining human affective responses to natural colorful images with multiple degree of emotional content. The data set contain 1182 images, each image were rated by humans with different genders, ages and cultures to have emotional values of valence, arousal and dominance in rage of (1,9). In our experiment, we consider of using only valence and arousal since two values show their influence to human cognition presented in Chapter 3 .

We select our input images based on the work of Lu [47] aimed at studying about shape features in evoking human emotions. In that, all images which contain human facial expression and body gestures, are removed because they strongly effect to human emotions based on body language which slightly changes can lead to evoke a completely different emotion. As a result, 484 images are selected and we divide our data set into two sets: a training set contains 80% of the selected images (388 images) and a test set consists 20% of the selected images (96) images. Some samples of selected images are presented in 2-dimension emotional framework is presented in Figure 4.9.

In addition, our input images are extracted in three channel of red, green and blue which are normalized based on standard score firstly before being trained by our network. To normalize our data by standard score, we use the formula given by:

$$X_{new} = \frac{X_{old} - \mu}{\sigma} \tag{4.3}$$

Here,

- μ is the mean of the data.
- σ is the standard deviation of the data.

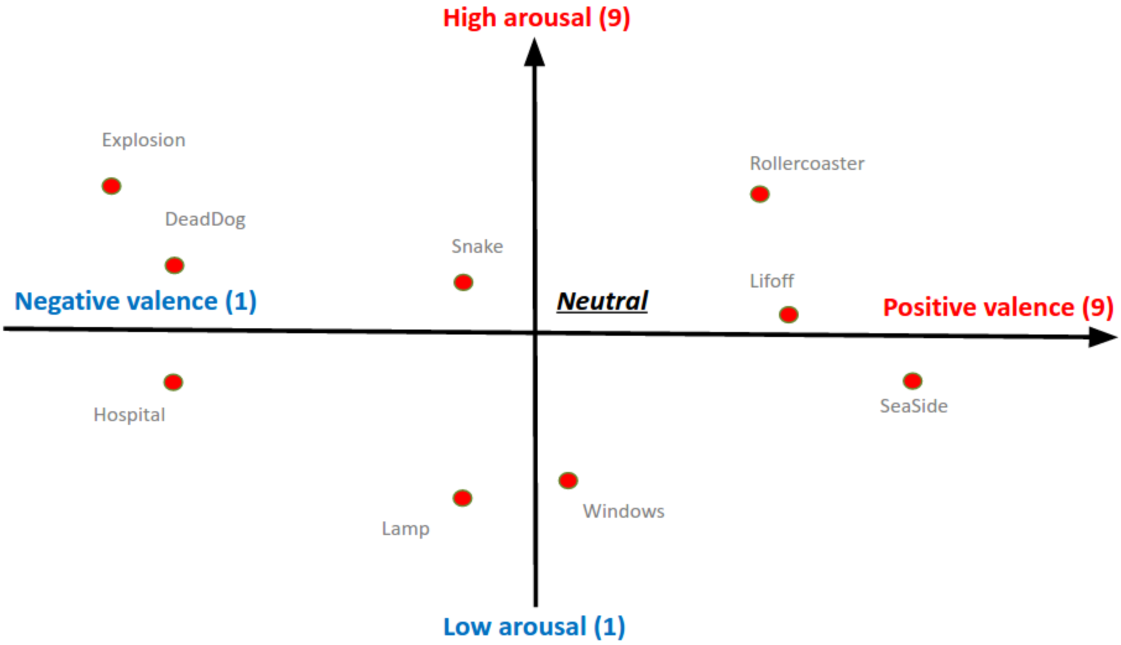


Figure 4.9: Some sample of selected data for the designed CNN is presented on the valence-arousal coordinate. Valence and arousal components of rated emotions for each image are ranged from 1 to 9.

Evaluation Method

We use Mean Squared Error (MSE) method to evaluate the performance of our designed network on generating valence and arousal values based on input images. MSE is used to compare the different between the output of the network and the expected output for the test set which contains 20% ogrinal data without being trained by our network through Equation 4.4 as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2 \tag{4.4}$$

Where,

- n is the number of samples in test data set
- Y is the output of the designed network at output layer
- y is the desired output which is obtained from the IAPS data set

Result of Acquiring Guide Emotion based on CNN

The result of our experiment is shown in Figure 4.10 in compared with Lu's work. Blue columns show our result while orange columns present for the result of Lu's work. All the result are measured by mean squared error. It is easy to claim that Lu's work obtained better results than our work. However, the different between the error on arousal dimension of our work in compared with Lu's work is smaller than the corresponding result on valence dimension. Besides, CNN is wildly known for working effectively with a big training data which can includes more than a million images [36]. In our experiment, we use only a small number of input data which included 488 images. In the future, we can use a bigger number of training images to improve the performance of our designed CNN. In addition, there are a lot of available accessing

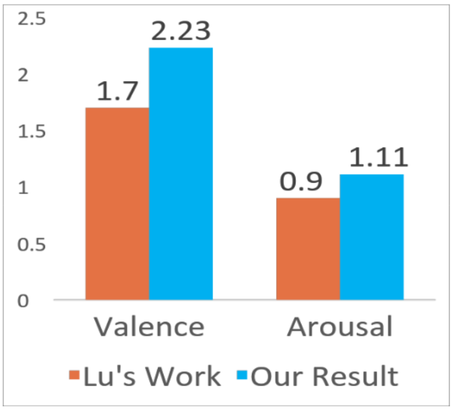


Figure 4.10: Mean Squared Error for valence and arousal dimensions of our works in compared with Lu's work.

pre-trained networks [36] [70], we can apply and fine-tune pre-trained networks for getting higher percentage of accuracy when applying CNN for this task. Therefore, this approach can be studied more to improve the performance of CNN on generating guide emotions from humans for robots.

4.2.2 Gaining Guide Emotion through HRI

Motivation

We designed a convolutional neural network to extract visual features automatically and thereby generating emotional values for input images. The result is considerable to be improved in the future when applying fine-tuning networks, available initial weights and a large number of input images. Proposed machine learning techniques recently can map one image to only one emotion which consists one combination of valence and arousal values.

However, in attempt of modeling human-inspired social effects for robots, we expect our robots can get multiple guide emotions from humans for the same image because for a certain image, different humans can have different emotions due to their background or culture, and even for one human in different times, the human can generate different emotions for the stimulus. In another word, we expect to map one image to multiple different emotions, which recent applied machine learning techniques cannot support. Thus, we propose to model human guide emotion through providing robots with those emotions directly from humans through HRI.

Implementation

As presented in Section 3.1, social effects play an important role in human development through social referencing and social sharing. It not only supports human infants to obtain basic knowledge about a particular stimulus or object for forming appraisals for future recalls but also allows older infants, children and adult to acquire a more detailed interpretation about the environment with social context. Furthermore, social effects also helps to facilitate human relationships during interactions and continuously learning and adaptation processes. Infants directly mimic caregiver and parent guide when being influenced by referencing perspective of social effects. While parents can make children feel comfortable to listen to parent guide and share with them about their personal experiences, strangers make children be silence or lie when asking about personal experience of children due to both aspects referencing and sharing of social effects. In addition, children 's knowledge contains more social context with deep understanding about complex objects, events and situations than infants'.

To model social effects to be given to robots, we consider human guides for the role of teaching and sharing during human-robot interactions by providing each robot with guide knowledge and the influence from culture. In other words, the guide provides each robot with the interpretation and emotions about a specific object so that robots can further create appraisals related to that object. Due to the influence of social referencing in infants who is younger than 18-month old, we model to make our robots play a role as an infant to interact with humans at initial state. In other words, our robot long-term memory is initialized emptily. After several interactions, robots can play a role as similar as human children and adult for reducing the agreement on human guide emotions as an affective factor for sharing knowledge about object appraisals and generating emotions. This influence depends on the intensity of the relationship between human and robots and guide emotion components. In our research, we suppose that all robots have the same relationship with humans, so the relationship between humans and robots are the same, only the intensity of guide emotions has influence on robot emotion generation process.

4.3 Designing Robot Long-term Memory for Acquiring Emotional Experiences

4.3.1 Motivation

Cognitive memory plays an important role in the development of social robots for actively learning knowledge and autonomously responding to the environment through past experiences and interactions with the environment. In cognitive robotic field, there are two main kinds of memory architectures which are cognitive memory architectures and developmental memory architectures. Cognitive memory architectures focus on modeling some aspect of human memory such as how information is stored and used for future retrieval. Thus, a lot of robot memory architectures were proposed to mimic very general functional processes and knowledge representation structures of a human memory such as Episodic Memory [18], Procedural Memory, and other integration of memory components [31]. This kind of architecture aims at simulating some specific cognition of humans to enable robots to finish some given tasks. In contrast, developmental memory architectures investigate the role of all memory components and the interconnection among them; in that, the interconnection plays a key role to form personality, acquire knowledge continuously and develop intellectually in general though a psychological-inspired epigenetic approach. Therefore, this approach can help robots to adapt their obtained knowledge to learn new fact through interactions with humans and the environment[78].

Based on the epigenetic approach, several memory architectures have been proposed [10] [50] [61]. Presented memory architectures of [10] [50] focused on individual memory components, ignored the role of context information in robot knowledge acquiring processes; thereby, robots react uniformly while humans have different responses to the same stimulus or situation. After a short-term of interactions, robots loose human attentions, which is called human-robot interaction gap. In attempt to reduce the gap, Epigenetic Robot Intelligent System (ERIS) [61] which was presented and implemented in a commercial robot, covered all the missing elements of previous epigenetic memory architectures and investigate the interconnection among those elements to enable robots to understand contextual information. The Long-term Memory (LTM) of ERIS is designed to have Semantic Memory (SM), Episodic Memory (EM), Procedural Memory(PM) and the interconnection among those components as presented in Figure 4.11. While SM provides robots with a container for generating general knowledge about objects or facts, PM acts as a passive component to store knowledge about robot behaviors such as motor skills. The link between them shows the connection of actions related to a certain

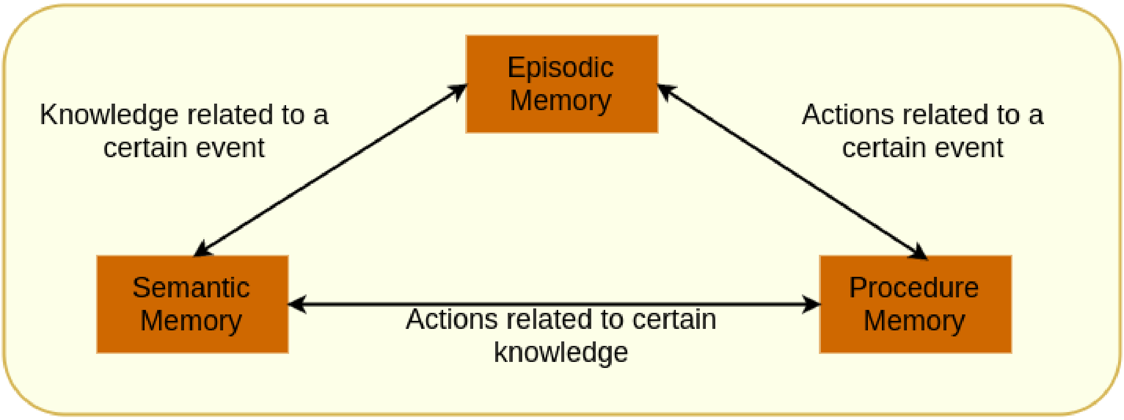


Figure 4.11: The interconnection of Long-term Memory components in terms of past experiences based on ERIS [61].

knowledge such as looking at the sun. Look is an action which is attached with the knowledge about the sun. The whole information of this event is stored in EM through the interconnection of SM with EM and PM with EM. As presented in the example, the knowledge about the sun as facts is consolidated in SM while the action look is maintained in PM to specify how a physical body can perform this action. The knowledge is connected and consolidated as an episode in EM to show other contextual information in general such as color of the sun at the time of looking. When consolidated episodes in EM are recalled through a retrieval process, all related elements of the experience to let a subject relive this experience due to the interconnection among memory components. Subjects can relive the color of the sun or the action looking when recalling all consolidated episodes in EM related to the experience. Thus, in our robot emotion representation, we adopt ERIS to design our robot long-term memory for enabling robots to encode, consolidate and retrieve experiences including emotions which are not included in ERIS; thereby using experienced emotions as a factor which influences robot emotion generation in respond to a certain stimulus from the environment during interactions with humans.

4.3.2 Object Affective Appraisal Formation

Human Object Affective Appraisals to Situation Awareness Capacity

Psychologists claimed that humans memory are sensitive to colors, shapes and locations of objects in order to recall past experiences and determine the similarity among objects [12] [40]. Due to color red, humans may recall several different experienced objects which are red such as roses, books and stones when they see a red wall. In addition, humans can also use the shape of an object to reduce the number of similar objects to a certain object. For instance, a red square can make children think about a red cube instead of a red rose. Color and shape of objects can evoke human emotions such as red color can make some people feel sad when thinking of blood, or being relax for watching roses otherwise. Some people have no problem of remembering shape of objects while others are fear of holes.

According to OCC emotion theory [55], human emotion generation is based on three main aspects: occurred events, actions of self-agent or others and interest on certain property of objects. In that, to understand the context of occurred events, humans need to have basic understanding about objects included in an event, personal interest on certain property of objects and the influence from people around due to the functionality of past experiences, emotion generation processes and social effects respectively. In infants, while their knowledge is directed and guided, their emotions and emotional expression are shaped by their parents due to social referencing which is mentioned in Section 3.1.1. Infants, then form their own basic affective appraisal about novel objects to make their own evaluation for negotiating with their parent guides on obtaining new knowledge, forming personality and generating emotions as the result of social sharing. After a long-term of learning and development, children and adults are able to form their own affective appraisals and interest about objects to make analysis for discriminating objects, interpreting a specific situation and thereby generating emotions and reactions.

Links among memory components make experiences obtained by a subject are related to personal emotion and emotional responses of the subject to a specific object or stimulus from the environment [41]. The subject may attach a specific emotion to a certain object and relive this emotion whenever an experience about this object is recalled. For instance, a child may keep facing his fear emotion for a ghost because his past experiences about ghosts are recalled whenever others tell him stories about ghosts. If the arousal intensity level of the environment increases, the child can face to a higher level of negative effects from recollected experiences as a result of emotion influence in memory enhancement presented in Section 3.2.

Modeling object-based Memory for Robots

In order to understand any stimulus or situation from the environment during interactions with humans and generate responded emotions, robots need to have basic knowledge about objects. The basic knowledge can be related to object key features such as color and shape due to their effect on human memory performance enhancement. Robots can use those features to form affective appraisals for comparing and evaluating unknown objects with experienced objects for a better interpretation about the environment. Robot long-term memory can be used to encode and maintain experiences continuously during interactions of robots with humans and the environment. For instance, robots are able to recognize an experienced object when seeing it at the current time in different locations. A robot may be given a scene of a yellow tennis ball and a red cube, this robot then form appraisals about the ball based on key features of the ball such as yellow, circle and soft. For the next interaction, the robot saw a white circle ball and asked to show the most relevant experienced object to the toy. The robot should answer the tennis ball instead of the cube.

Besides, the interconnection among memory components can provide robots with a capacity for acquiring not only knowledge about objects but also interpretation about related situations effectively. ERIS was proposed to give robots a developmental memory architecture which considers fully the link among memory components including semantic memory and episodic memory. The information about objects is stored in semantic memory which links with related episode stored in episodic memory for a certain experience. Robots can use the knowledge to compare and evaluate unknown objects with known objects to have more appropriated understanding about a new experience at the current time. For example, robots can detect and localize an object in different locations at different times when the object is moving due to several captured images; thereby re-locate the original location of this object.

4.3.3 Memory Error in Re-consolidation Phase

Memory re-consolidation in humans refers to the concept that an experience after recalling because of the activation from similar experiences, is updated in LTM. It means that the episode and related semantic items corresponding to the experiences are more possible to be accessed in the future. There are two main different versions of memory re-consolidation. The first one emphasized at that while a lot of cognition processes happen in human brains, the consolidated information is unchanged and a passage of time corresponding to the experience is merged during memory re-consolidation phase [71]. In contrast, Multiple Trace Theory (MTT) aimed at that re-consolidation processes of recalled experiences make the change for the original episodes and engaged semantics items via a trace [52]. When an experience is consolidated in LTM, a trace is created. Whenever this episode is recollected through memory retrieval, a consolidation process is required and the trace is updated due to the change of this episode. This consolidation process may cause memory error which an episode is re-consolidated to be totally different from the original one or this episode is lost. It leads to form memory disorder. Besides, a new episode including a new trace may be created based on a part of the old episode such as semantic items and the trace of the old episode is updated due to this change. The higher number of accessing time of the episode increases the possibility of successful recalling this episode in the future.

To model memory re-consolidation process for our robots, we apply MMT to create a trace for every episode and update the trace of an episode whenever it is recalled. Besides, this theory also supports that humans encode information partially which may help them remember some specific characteristics of an object without remembering the object itself. Therefore, it satisfy our goal of forming object affective appraisals and acquiring knowledge due to those appraisals. Although this process provides robots with ability to acquire new knowledge and create the link between their personal experiences to new interpretation about similar stimuli and objects from the environment, it can cause memory disorder. In order to eliminate the possibility of creating memory error and memory disorder, we only update traces of episodes without changing the original episodes.

4.3.4 Implementation

Our Long-term Memory (LTM) is designed based on two main components SM and EM in order to enable robots to interpret context information and obtain knowledge intellectually. Due to the adopted memory architecture, EISR, we study the interconnection among EM and SM. Robot SM acts to encode basic interpretation about a particular object and form appraisals based on the knowledge. Thus, robots are able to distinguish multiple objects due to their key features for future usage. EM is designed to consolidate the experience of robots including engaged emotions related to objects or stimuli from the environment. For example, robots are able to remember the position of objects, engaged emotions and number of objects at a certain time. Due to the link between SM and EM, robots can use formed appraisals to evaluate a new object based on recalled past experiences which contain similar known objects through memory processes.

In order to perform memory processes for acquiring robot personal knowledge and generating emotions, we adopt several definitions which were presented in ERIS, and propose new definitions as following:

- **Definition 1: A scene**, a raw image, is captured by robots or given by humans at a certain moment during human-robot interactions. This scene is encoded by robots based on human guide to be consolidated in robot LTM for future usage. Furthermore, all scenes have to be uniform at resolution and size which can be used easily for other processes. From multiple different scenes of one object, a robot can measure the distance between the object and the robot such as near when the size of the object is small and far otherwise.

- **Definition 2: A time-stamp** is the time robots capture a scene; thereby attaching with the encoded episode to be maintained in LTM. This time-stamp is used for evaluating the age of an experience of knowledge which robots acquired.

- **Definition 3: An object** is a specific stimulus such as cube which is contained inside a scene. Robots can use human guides through interactions or obtained knowledge after several interactions with humans or the environment to generate appraisals about this object. The knowledge is encoded and consolidated in robot SM.

- **Definition 4: An object location** is a minimum bounding box which contains the detected objects. This area can be identified by robots or guided by humans through interactions. From the identified area, robots can measure the height of the object and localize the object within a scene. For examples, robot can make a comparison between two cube insides a scene to get information of the bigger cube due to the larger area which this cube takes place.

- **Definition 5: A semantic item** is a representation of robot basic interpretation about an object such as color or shape after processing a scene at consolidation phase of memory. This item is stored in SM and linked by several episodes due to the interconnection of EM and SM. A corresponding item of an object can be used to form affective appraisals for making prediction for unknown objects through memory retrieval phase.

- **Definition 6: Engaged emotion** of a scene is the robot personal emotion at the time of consolidating the scene. This emotion is generated after updating robot personal emotion, and attached to a certain episode to present the emotion in respond to the scene.

- **Definition 7: An episode** which is a representation of a scene in robot EM after memory consolidating phase and emotion generation processes, contains links to one or several semantic items which describe contained objects in the scene.

- **Definition 8: A trace** is a attached information about the encoded time of the episode in respond to a specific stimuli. Whenever the episode is recalled, the trace is updated by adding an information of retrieval time.

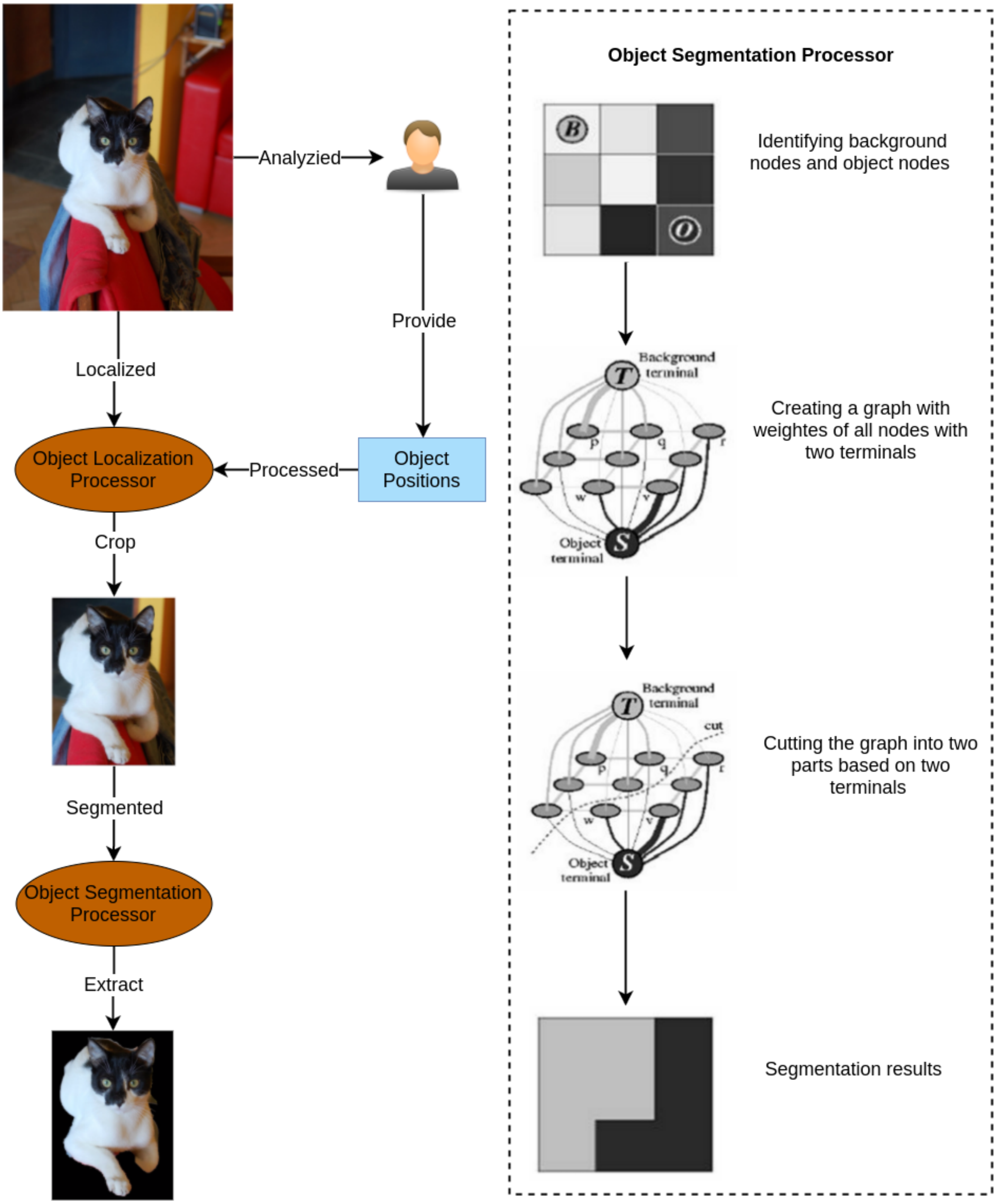
In addition, for ethical reason and memory performance, we design and experiment our robot long-term memory based on several conditions:

- **Assumption 1:** At initial state, every robot has no prior knowledge; in other words, robot long-term memory is empty with no data. Robots are able to use personal knowledge right after acquiring from the first interaction to update personal emotions, obtain a more detailed knowledge and possibly form personality.

- **Assumption 2:** There is no forgetting mechanism is applied because this mechanism can cause memory errors. In other words, our robots are able to encode all experiences and use those consolidated knowledge for future recall with one hundred percentage of accuracy.

Consolidation Process

During interactions with humans, robots can generate scenes via a provided camera or given images and get human guided information about a certain stimulus. For a certain amount of time which is set by default, robots are able to capture images during interactions with humans and the environment. Robots can get raw images from humans directly through interactions. All captured images are re-sized to a standard resolution to become scenes for localizing objects and extracting features of those objects. In additions, the provided information from humans contain guide emotions which help to shape robot emotion for more acceptable by humans and object localized areas and object labels which enable robots to acquire personal knowledge.



(a) Extracting a target object from a noise background image [48].

(b) Image segmentation process based on Grabcut algorithm [2] and color information.

Figure 4.12: Extracting an object from a captured image during human-robot interaction based on human guide and graph segmentation.

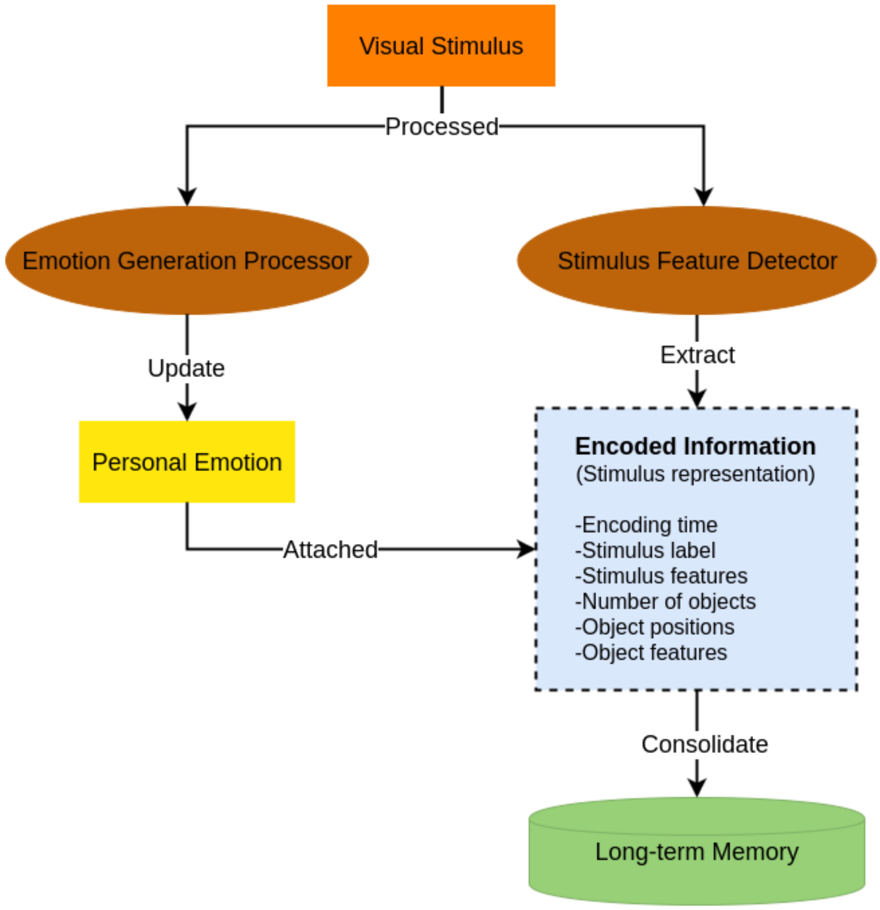


Figure 4.13: Memory consolidation process after encoding a certain visual stimulus and generating responded emotion to form an emotional experience consolidated in robot long-term memory for future retrieval.

```
Obj_abc123.sm = {
   label = abc123
   image = (RGB image)
   color = (HIST array)
   shape = (SIFT descriptors)
}
```

(a) An example of a semantic item consolidated in Semantic Memory.

```
Stimulus_xyz789.em = {
   encoded_time = (timestamp)
   label = xyz789
   obj_count = 1
   obj_position = [5,1,80,123]
   obj_name = abc123
   emotion = [2.121, 5.620]
   trace = def801
}
```

(b) An example of an episode consolidated in Episodic Memory

```
Trace_def801 {
   label = def801
   retrieval_count = 1
   retrieval_record= [1, (timestamp)]
}
```

(c) An example of a trace engaged with an episode.

Figure 4.14: Structure of semantic and episodic memory in robot Long-term Memory.

Based on the localized areas of objects and the scene, robots are provided with a function named Visual Feature Detector (VFD) for removing noise background and extracting object key features. There are two main steps which this function acts to remove background of the scene. Firstly, robots localize object locations based on provided information from humans to get minimum bounding box around the target object through Object Localization Processor sub-function of VFD. The capture image then is cropped to have the size which equals to the identified minimum bounding box. Lastly, robots extract the target object by replacing the noise background by a black background through Grabcut algorithm [64] attached in Object Segmentation Processor sub-function of VFE. This sub-function aims at making a weight graph of the image. The graph which links pixels due to their color information, has two roots to present for the background terminal and the target object. The similarity between the color of the root pixel and neighbor pixels is presented by the weight between nodes in the graph. After getting all weights of links among the graph, the background of the image can be separated from the target object. This process is presented in Figure 4.12.

In order to form an experience and object affective appraisals for a certain stimulus, VFD function helps robots to extract visual features of the object after removing noise backgrounds. Basic visual features of the detected object: color and key points due to Section 4.3.2 are extracted by using color histogram (HIST), and scale-invariant feature transform (SIFT) [46]. Those features are encoded to become a visual package of a semantic item which can form robot personal appraisals in order to recall past experiences for acquiring new knowledge and updating robot personal emotion in respond to the scene. On the one hand, semantic items corresponding to all detected objects in the scene are linked to form a new episode. On the other hand, those items can recall past experiences of robots and allow robots to relive experienced emotions; thereby changing robot personal emotion due to the new interpretation and Emotion Generation Processor (EGP) which is described in the next chapter. The updated emotion is attached with encoded episode for readily consolidating in robot LTM for future recall. Memory consolidation process is shown in Figure 4.13, and consolidated memory items in LTM is illustrated in Figure 4.14.

After consolidating the new experience in robot LTM, robots maintain episodes and related semantic items with the link among those components for future use. Those links can help robots to recall past experiences for evaluating new stimuli and form appraisal for acquiring knowledge from those stimuli through memory retrieval phase.

Retrieval Process

When robots process a new experience, they can evaluate the similarity between new objects and with seen objects via formed affective appraisals. As mentioned in Section 4.3.4, robots can encoded visual features of a new object and form appraisals from those features; therefore, using those appraisals to evaluate the new object with all experienced objects via the corresponding consolidated and maintained semantic items of experienced objects in LTM. We define the similarity between two objects by comparing object visual features: color and key points of the new object (O_{new}) to a previously seen object (O_{seen}). The Euclidean distance is used for the calculation of the difference between color (HIST) and key points (SIFT) of two objects, where the difference is represented as d_{color} and $d_{keypoint}$ respectively. The role of each difference on the distance calculation is defined as the parameter ϵ given by.

$$d_{(O_{new},O_{seen})} = \epsilon \cdot d_{color(O_{new},O_{seen})} + (1-\epsilon) \cdot d_{keypoint(O_{new},O_{seen})} \tag{4.5}$$

Figure 4.15: Example of images [48] contain objects which are similar in color features.

Algorithm 1: Retrieval of the most similar experienced object in compared with the unknown object

Input : SemanticMemorydirectory, $SIFT_{newObject}$, $HIST_{newObject}$, ϵ
Output: similarObjectLabel
similarObjectLabel = null ;
if *!SemanticMemorydirectory* **then**
  listObject = readAllFile(SemanticMemorydirectory);
  numObject = length(listObject) ;
  i = 0 ;
  count = numObject;
  listKeypointDistance = empty ;
  listColorDistance = empty ;
  listTotalDistance = empty;
  while *count > 0* **do**
    $SIFT_i$ = listObject[i].getSIFT() ;
    $HIST_i$ = listObject[i].getHIST() ;
    keypointDistance = math.Euclidiandistance($SIFT_{newObject}$, $SIFT_i$) ;
    listKeypointDistance.insert(keypointDistance) ;
    colorDistance = math.Euclidiandistance($HIST_{newObject}$, $HIST_i$);
    listColorDistance.insert(colorDistance);
    count- - ;
  end
  maxKeypointDistance = max(listKeypointDistance) ;
  maxColorDistance = max(listColorDistance) ;
  for *i <= numObject* **do**
    CDrate = $\frac{colorDistance[i]}{maxKeypointDistance}$;
    KPDrate = $\frac{keypointDistance[i]}{maxKeypointDistance}$;
    totalDistance = $\epsilon \cdot CDrate + (1 - \epsilon) \cdot KPDrate$;
    listTotalDistance.insert(totalDistance) ;
  end
  minIndex = getIndex(min(listTotalDistance)) ;
  similarObjectLabel = listObject[minIndex].getLabel();
  i++ ;
end
return similarObjectLabel ;

Algorithm 2: Relive experienced emotion based on an object appraisal

```
Input   : EpisodicMemorydirectory, objectLabel
Output: emotion
valence = 0 ;
arousal = 0 ;
emotion = [valence,arousal] ;
listFoundEmotion = empty ;
listFoundTrace = empty ;
listAccessingTime = empty ;
if !EpisodicMemorydirectory then
    listEpisode = readAllEpisode(EpisodicMemorydirectory);
    numEpisode = length(listEpisode) ;
    numTrace = length(listTrace) ;
    i = 0 ;
    for i < numEpisode do
        objectLabelInEpisode = listEpisode[i].getObjectLabel();
        if objectLabelInEpisode = objectLabel then
            valence = listEpisode[i].getValence();
            arousal = listEpisode[i].getArousal();
            emotionTmp[0] = valence ;
            emotionTmp[1] = arousal ;
            traceLabelInEpisode = listEpisode[i].getTraceLabel() ;
            trace = getTraceByName(traceLabelInEpisode) ;
            accessingTime = getLastAccess(trace) ;
            listFoundEmotion.insert(emotionTmp) ;
            listFoundTrace.insert(traceLabelInEpisode) ;
            listAccessingTime.insert(accessingTime)
        end
        i- - ;
    end
    indexOfRecentAccess = argmin(listAccessingTime) ;
    emotion = listFoundEmotion[recentIndex] ;
    traceNeedToUpdate = listFoundTrace[recentIndex] ;
    timeStamp = Time.now();
    traceNeedToUpdate.insertRecordOfAccessing(timeStamp);
end
return emotion ;
```

Due to the calculated distance, we get the smallest distance which shows the most similar object previously seen to the new object. Some examples of similarity recognition of our robots are shown in Figure 4.15. Due to the semantic item corresponding to a seen object and links from the item to episodes in EM, we can find all episodes which contains the item through memory retrieval process. In case of having only one object inside an image, only the episode contains the similar semantic item is recalled. If there are more than one object insides an image, the number of objects are used to determine which episode is recalled. In case of getting multiple similar episodes, attached time-stamps of those episodes are used to get the most recent experience for retrieval. When the experience is recalled, the related episode is updated by adding retrieval time to the engaged trace of this episode, and robot can relive personal past experiences based on the information from the recalled episode including experienced emotions and other contextual information. For obtaining experienced emotion based on memory retrieval for visual stimuli contain only one object, we designed algorithms 1 and 2 based on Equation 4.5 and the interconnection between EM and SM in our robot LTM.

Limitations

Our robots need to have human guide on object areas within a scene to be able to extract visual features of objects. This function is developed imperfectly, our robots cannot detect objects and localize positions of detected objects within a scene. It lead to narrow the ability of robots to develop independently from humans. Besides, the Grabcut algorithm sometimes works not well on removing complex backgrounds from the localized object areas due to the quality of raw images, the color of object which is similar to background color. In our experiment, we use images which have simple background or their backgrounds have contrast colors with colors of the target objects.

For the similarity of objects through object affective appraisals, we consider color histogram and SIFT features which are provided by OpenCV library. Those algorithm do not work well for complex stimuli such as living organism from the environment. Robots cannot recognize the same dog while having multiple scenes of different dog postures. Besides, robots are not able to distinguish different kinds of animal and breeds of them. In our experiments, we use multiple images of only one category, cat, from available international data sets and we expect our robots can recognize similarity between two cats through their postures such as lying and their hair colors such as black and white.

4.4 Integrating Multiple Cues for Representing Robot Emotion

In our robots emotion model, we use visual stimulus as the factor to evoke robot personal emotions. Robot personal emotions changes due to time parameter during human-robot interactions or the exploration of robots to the environment. Robots acquire new knowledge from human guides and generate emotions in respond to the new interpretation for continuously adapting to the environment and developing in general. For each visual stimulus captured from the environment during interactions with humans, our robots are expected to form appraisals about objects, obtain personal knowledge and update personal emotions through the influence from human-inspired emotion activation cues. Robots are expected to play two roles which are a learner and a partner to continuously learn from and share with humans.

As presented in Section 4.2, our robots can get human guide emotions directly through HRI due to the influence of social effects. For the same stimulus, a robot can obtain different guide emotions from different humans based on their background and culture. In other words, we enable our robots to be influenced by culture and supervised by humans through interactions to generate emotions and acquire knowledge about the environment. Thus, human-inspired social effects are modeled as guide emotions which are given directly by humans during interactions with robots.

The influence of past experiences in human emotion generation is modeled as experienced emotions which consolidated with a certain knowledge and maintained in robot LTM as shown in Section 3.2. Our robots are able to acquire personal knowledge based on the supervise from humans to form basic object appraisals which can be recalled in the future when robots are affected by similar stimuli. In addition, every experiences of robots are engaged with a certain emotion which can be neutral, happy or sad represented in the valence-arousal coordinate. Those experienced emotions can be evoked with one hundred percentage of accuracy due to

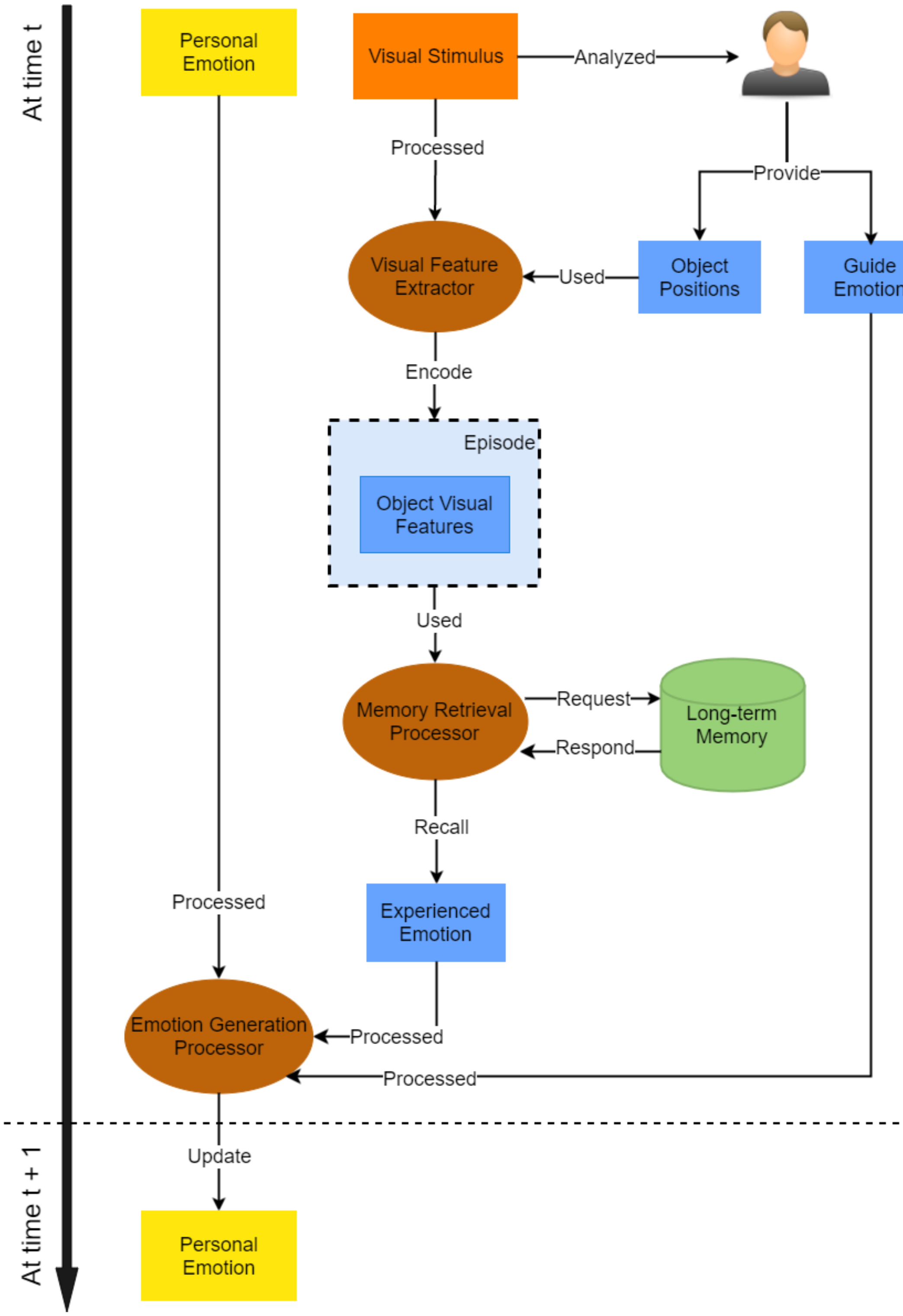


Figure 4.16: Robot emotion generation process based on human-inspired social effects and robot personal experiences through memory retrieval for a certain visual stimulus.

our assumption in Section 34 when robots recall their past experiences which contain those emotions emotions. We consider that there are experienced emotions can impact to the current emotional state of a robot, which plays similar role as past experiences in humans on activating emotions in respond to a certain stimulus, object or situation.

Table 4.1: Denoted symbols for multiple human-inspired emotion activation cues in robot emotion generation.

Symbol	Meaning
C_{GE_t}	An emotional component of the personal emotion of a robot at the time encoding the stimulus (t)
C_{PE_t}	An emotional component of the personal emotion of a robot at the time encoding the stimulus (t)
C_{EE_t}	An emotional component of the experienced emotion from a human at the time encoding the stimulus(t)
$C_{PE_{t+1}}$	An emotional component of the personal emotion of a robot after reacting to the stimulus ($t+1$)
γ_1	The impact of C_{GE_t} to $C_{PE_{t+1}}$ due to the intensity level of C_{GE_t}
γ_2	The impact of C_{EE_t} to $C_{PE_{t+1}}$ due to the intensity level of C_{EE_t}
γ_3	The impact of C_{PE_t} to $C_{PE_{t+1}}$ due to the intensity level of C_{PE_t}

Our robot personal emotions changes due to time parameter during human-robot interactions. Robots acquire new knowledge from human guides and generate emotions in respond to the new interpretation for continuous adaptation and to the environment and development in general. For each visual stimulus captured from the environment during interactions with humans, our robots are expected to generate appraisals about objects, obtain personal knowledge and updated personal emotions for both role of learner and sharer. The overview of this integration is shown in Figure 4.16.

When interacting with humans, robots are provided with the location of all objects in a visual stimulus and a guide emotion. On the one hand, the provided object location helps robots to extract object visual features. Those feature can be used for generating a new episode and semantic items through memory consolidation phase and forming object affective appraisals as presented in Section 4.3.4. Those appraisals help robots to recall past experiences including engaged emotions through memory retrieval phase described in Section 4.3.4. The recalled emotions can influence to robot personal emotion in respond to affective stimuli. All the processes in the memory retrieval phase are compressed and performed by Memory Retrieval Processor function which sends request to robot LTM and receives responses from the memory. On the other hand, the guide emotion shapes robot personal emotion to be more favorably accepted by humans. Moreover, Emotion Generation Processor (EGP) function is designed to update robot personal emotion based on the guide emotion and an experienced emotion. This function helps robots to analyze the influence of human guide as social effects in humans and robot personal knowledge as the role of past experiences for generating robot personal emotion in respond to a certain stimulus from the environment. In order to update robot personal emotion, we denote several symbols presented in Table 4.1.

The personal emotion of the robot at time $t+1$ is created by the personal emotion of the robot, guide emotion and experience emotion the robot acquires at time t and the influence of those emotions due to their intensity levels. Thus, the personal emotion of the robot at $t+1$ is calculated as Equation 4.6:

$$C_{PE_{t+1}} = \gamma_1 \cdot C_{GE_t} + \gamma_2 \cdot C_{EE_t} + \gamma_3 \cdot C_{PE_t} \tag{4.6}$$

Since the total influence of three emotions C_{PE_t}, C_{GE_t} and C_{EE_t} to $C_{PE_{t+1}}$ is calculated by Equation 4.7 given by:

$$\gamma_1 + \gamma_2 + \gamma_3 = 1 \tag{4.7}$$

Where γ_1, γ_2 and γ_3 denotes for the influence of the affective emotions to the personal emotion of the robot at $t+1$ as presented in Table 4.1 are calculated as follows:

$$\gamma_1 = \frac{C_{GE_t}}{C_{GE_t} + C_{EE_t} + C_{PE_t}} \tag{4.8}$$

$$\gamma_2 = \frac{C_{EE_t}}{C_{GE_t} + C_{EE_t} + C_{PE_t}} \tag{4.9}$$

$$\gamma_3 = \frac{C_{PE_t}}{C_{GE_t} + C_{EE_t} + C_{PE_t}} \tag{4.10}$$

Based on Equation 4.6 and 4.7, our robot personal emotion can be calculated by Equation where γ_1 and γ_2 are described in Equation 4.8 and 4.9 respectively.

$$C_{PE_{t+1}} = \gamma_1 \cdot C_{GE_t} + \gamma_2 \cdot C_{EE_t} + (1 - \gamma_1 - \gamma_2) \cdot C_{PE_t} \tag{4.11}$$

In addition, our robot emotion representation based on 2-dimension emotional framework as described in Section 2.2.2 which shows robot emotions based on a combination of valence and arousal values. However, two components have different influence on the subject for both emotion activation cues which are past experiences and social effects. Through memory processes, both negative valence and positive valence shows their effects while only high arousal does the similar influence to memory performance enhancement. In other words, valence and arousal act differently because of their intensity levels. In addition, according to Section 3.1, valence and arousal impacts in sharing aspect of social effects based on the same mechanism as influencing on memory performance. We assume that those elements play a similar role in referencing perspective of social effects since infants mimic their parent guide as a passive learner. Thus, we use a parameter, μ, to present the smallest intensity level of each emotional component and update Equation 4.8 and 4.9 respectively given by:

$$\gamma_1 = \frac{|C_{GE_t} - \mu|}{|C_{GE_t} - \mu| + |C_{EE_t} - \mu| + |C_{PE_t} - \mu|} \tag{4.12}$$

$$\gamma_2 = \frac{|C_{EE_t} - \mu|}{|C_{GE_t} - \mu| + |C_{EE_t} - \mu| + |C_{PE_t} - \mu|} \tag{4.13}$$

After robots update personal emotion, the generated semantic items and episode is combined with the generated emotions in respond to the visual stimulus to be consolidated in LTM for the complement of memory consolidation phase which is presented in Section 4.3.4.

4.5 Conclusion

We investigated the role of social effects and past experiences in robots to update robot personal emotion through the integration of guide emotion and experienced emotions. Guide emotions are provided by humans directly through interactions with humans due to the influence of culture. The influence enables robots to have various different guide for the same stimulus to acquire knowledge and generate emotions with context of culture, which are not provided by recent machine learning techniques. We also designed our robot Long-term Memory based on a human-inspired developmental memory architecture. The memory helps robots to form affective appraisals for stimuli in the environment for generating emotions and obtaining personal knowledge. Emotional experiences of robots are consolidated in the memory and recalled through HRI when robots are affected by stimuli from the environment which are as similar as previous experienced stimuli. When recalling previous experiences, retrieved emotions can let robots face to the same effect as the time of encoding those experiences. The guide emotion, the experienced emotion and the personal emotion of robot at the time of encoding the new stimulus influence to generate robot emotion in respond to that stimulus. We propose to apply an approach based on the percentage of the influence of each affective emotion to generate robot emotion.

Chapter 5

Experiments

In order to analyze how a system works, experiments and measurement processes are evaluated. In our research, we analyze the performance of our robot emotion representation by designing two main experiments due to a proposed scenario. We expect our robots can perform a role as not only a passive learner but also an active partner during interactions with humans. Besides, the adopted long-term memory architecture enables robots to obtain knowledge for future usage. Robots are expected to form personal evaluation on every stimulus from the environment based on visual images and human guide. We selected input data based on official international well-known databases for our robots to process and generate emotions. The generated emotions of robots are analyzed based on psychological experiments in humans. The details of experiment procedure and results are presented in following sections.

5.1 Data Selection

We set our experiments by running algorithms and evaluate generated emotions of robots in responds to a visual stimulus by analyzing the values of valence and arousal, thus we select images which are rated by humans to have those significant values. To model human guides during interaction with robots, we select our input data from international presented data sets of visual images which have been used for a lot of experiments in humans and applied for robots in psychology and cognitive science researches.

5.1.1 International Affective Picture System (IAPS)

IAPS [42] is the mostly wildly used data set of pictures which are used to study about emotions and attentions in psychology researches and robotic studies. This data set has developed and updated recently to contain 1182 natural color images in multiple different categories of human bodies, human faces, human gestures, animals, landscape,..etc. This data set is created and evaluated by humans based on multiple dimension theory of affects through values of valence, arousal and dominance. Those values are rated to have a standard deviation and mean in a range of 1 to 9. For each dimension, the lowest intensity level is presented by value 1 and in contrast, 9 shows the largest intensity level. The rating of IAPS is shown in Figure 5.1. IAPS has been applied to study about human emotions, how humans recall their past experiences through cognitive processes and brain activities [49], then evoking their emotions when seeing a specific image. For robotic field, biological signal from humans when watching images from this database is processed through several sensors attached with human skins during human-robot interactions to help robots to gather information such as heart beats, blood pressure, temperature or brain waves for recognizing human emotions and generating intelligent behaviors [62] or build social robots.

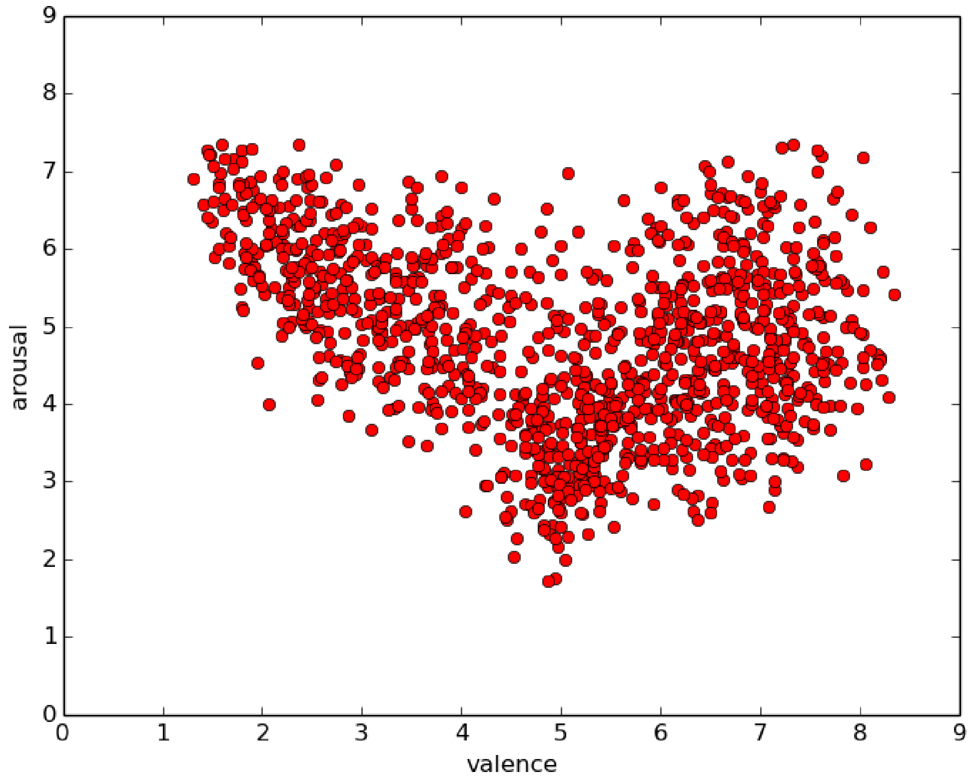


Figure 5.1: The presentation of IAPS database on the valence-arousal coordinate.

However, images in IAPS are different in quality and resolution, and the number of images in a specific content category is too small such as 4 images for cat; thus, it may lead to get noise in evaluating experimental results. Luminance, size and quality of images are claimed to have strongly effect to the subjects during experiments of evoking emotions and cognition processes especially when using different quality images in a categories [53]. In addition, unequal numbers of images in different categories and a small number of images in a category lead to present one image to a subject more than one time; thereby creating high effect to experimental results. For example, when a person is asked to see an image of a beautiful cat for the first time, he generate really high intense of emotion such as excited; however, for the second time of watching, the level of his emotion is reduced, he may feel happy. It is not easy to extract visual feature to study and recognize emotions from the data set due to the high number of image categories and the small number of images in each category even though machine learning technique is applied.

5.1.2 Nencki Affective Picture System (NAPS)

NAPS [48] database which is proposed recently to deal with the limitation of IAPS, contains a 1,356 realistic, high-quality images divided into five categories of people, faces, animals, objects, and landscapes. The number of images for each categories are balanced with the same resolution and quality. For each picture, rating on valence, arousal and avoidance-approach are created continuously due to a range of 1 to 9 by humans with a mean and standard deviation. NAPS is presented on the valence-arousal coordinate as shown in Figure 5.2. Furthermore, the information of some visual feature such as image contrast, image luminance and the number of green color pixels . This data set which provides higher number of images with lessen number of categories for high quality and resolution, is applied for multiple psychology researches for

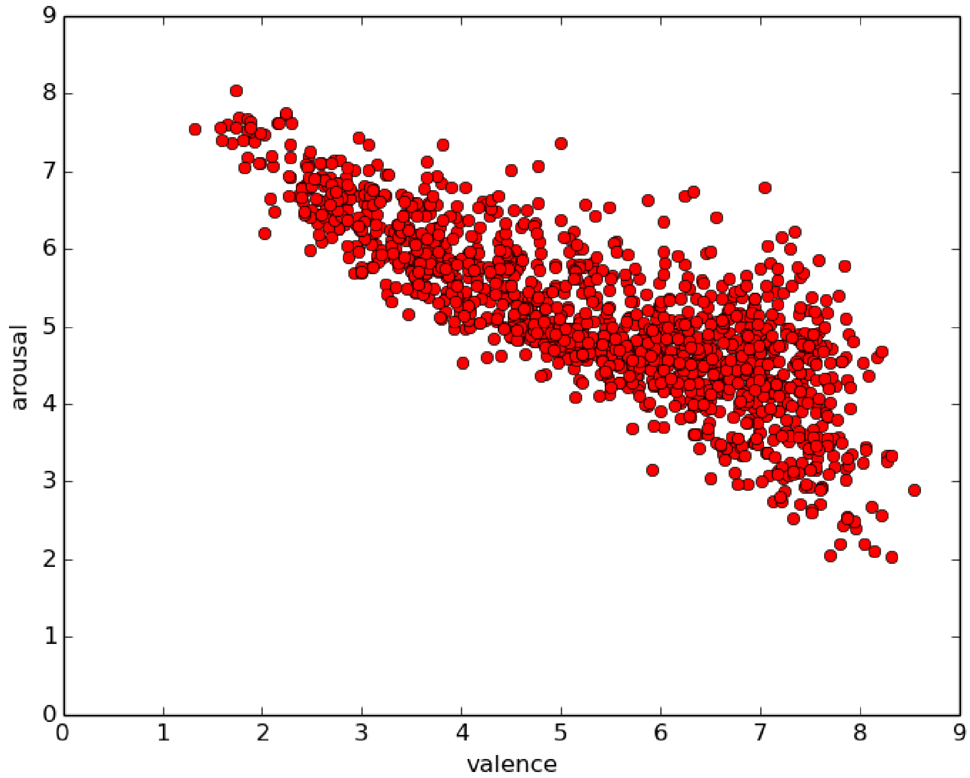


Figure 5.2: The presentation of NAPS database on the valence-arousal coordinate.

exploring human cognition, attention and emotions recently [59].

Although this database provides with more benefit than IAPS database, applying NAPS is still a challenging approach especially in robotic cognitive science. Robots do not have ability like humans to understand context meaning within an image then generating emotions and expressing reactions. Beside, a lot of images contain multiple objects with complex backgrounds; it leads to a difficult task of computer science when dealing with object detection, localization and recognition or situation awareness. Considering only simple background images, the number of images is reduced and unbalanced for each category. For example, to study about one category of animals, there are multiple kind of animals such as dog and cat. However, considering about kinds of animal to classify images, the number of re-categorized images is reduced extremely. It makes visual feature detection and processing tasks become more difficult due to the lacking of data.

5.1.3 Selected Data Inputs

In our research, we propose to represent robot emotions and process robot cognitive processes based on objects from visual stimuli while visual recognition tasks can be done by other researches. Thus, we only use visual stimuli with simple background and contains one kind of object. Besides, robots can localize objects within an image or scene based on human assists. Therefore, we assume that, our experiments are implemented to detect one kind of object which is cat. Due to the small number of images, we combine all images about cat in IAPS with all corresponding images in NAPS to have a bigger amount of visual stimulus which plays as an input in our experiments. After removing images which contains complex backgrounds and low quality, we got 12 images with valence (mean: 5.365, std: 3.525) and arousal (mean: 5.095, std: 2.405) when ignoring the role of dominance factor and avoidance-approach values in our robot

(a) A neutral-mid arousing image selected from NAPS [48].

(b) A positive-low arousing image selected from NAPS [48].

Figure 5.3: Samples of experimental input data selected from IAPS and NAPS databases for only one category of cat with simple background and contain only one object.

emotion representation. All images are re-scaled to the same size. Those images also contain only one object, so we assume that the rating of humans for each image is the rated emotion for the contained object. Because of unbalancing data and corresponding values of valence and arousal; we divided chosen images into four types such as negative-low arousing, negative-high arousing, positive-low arousing, and neutral-mid arousing. The example of selected images are shown in Figure 5.3.

5.2 Experimental Scenario and Procedure

In our experiments, we expected our robots to represent emotions by a combination form of two valence and arousal values. The robot may present not only influence of guide emotions when being asked but also display the past experienced stimulus and its effect. Besides, robots are able to discriminate different visual stimuli based on their visual features and forming object affective appraisals after encoding information about those stimuli.

In order to interpret how our system works, we designed a scenario in details, where humans provide guide emotions of every input image to robots. Images are presented to a robot in a certain sequence. In this scenario, the robot plays a role as a learner who can choose to accept human guide directly or negotiate with the guide based on personal gained experiences. In addition, the robot can show emotions through valence and arousal values. The scenarios is designed by three main steps. Firstly, a human gives the robot a visual stimulus and guide interpretation about the stimulus through emotional values and object locations. After getting the stimulus and human guide, the robot extracts visual feature of the image and included objects based on the information about given object locations. The encoded information is used for recalling robot past experiences and consolidating a new episode. When looking for similar experiences based on object visual features such as shape and color, the robot can relive experienced emotions which can influence to robot personal emotion. Secondly, the robot integrate all influenced emotional values with it own personal emotions to respond an emotion for the stimulus based on Equation 4.11. Lastly, after updating personal emotion due to internal and external information, the robot combines the generated emotion with encoded episodes to form new affective appraisals; thereby consolidating the obtained experience in robot LTM for future usage. All assumptions in our robot LTM design are applied due to Section 34.

For evaluating our robot emotion representation, we design two main experiments due to our proposed scenario. The first experiment aims at analyzing the influence of social effects and how our system work to model this influence to robots in cognitive processes for generating emotions. For this experiment, our robots play two role as a passive learner and an active partner to be influenced by social effects for both aspects of referencing and sharing when acquiring personal knowledge and generating emotions. We expect our robot to generate distinctive emotions for the same visual image if it observed at different time instants. The second experiment aims at evaluating the integration of multiple emotion activation cues in robot emotion generation inspired by human psychology theories. In this experiment, we expect two different robots generate different emotions for the same image at the same time of being affected due to their prior personal knowledge and experiences when applying the same emotion generation system and getting the same human guide. For both experiments, we provide robots with several images in a certain order, robots then display their personal emotions, guide emotions and experienced emotions when successfully obtaining similar past experiences from LTM.

5.3 Experiment 1: Investigating Human-inspired Social Effects in Robot Emotion Generation

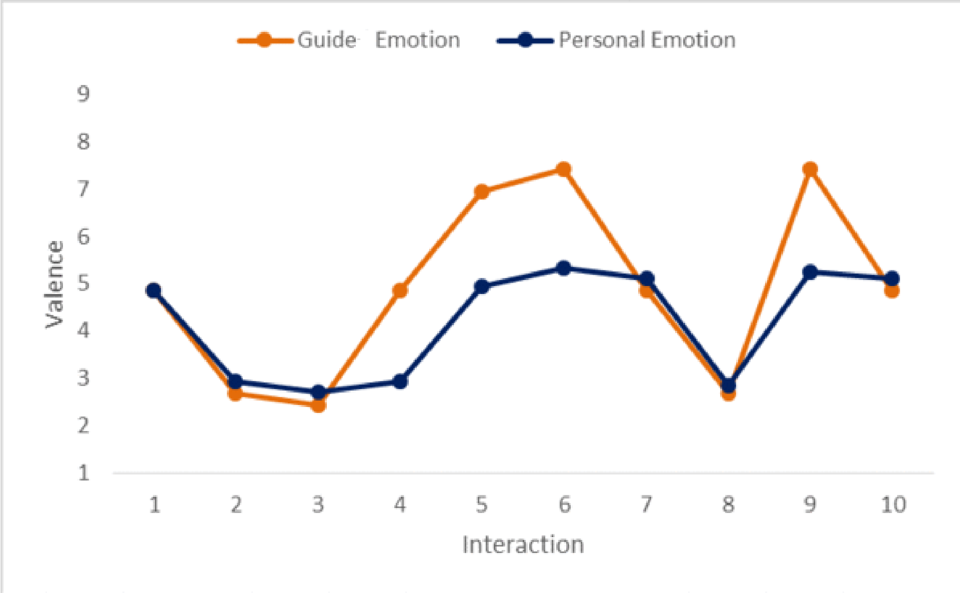
In this experiment, we expect our robots can play two roles which are a passive learner and an active partner to learn from and share with humans. Due to provided information from human, robots can acquire knowledge continuously and generate emotions automatically in order to adapt to the environment.

5.3.1 Design

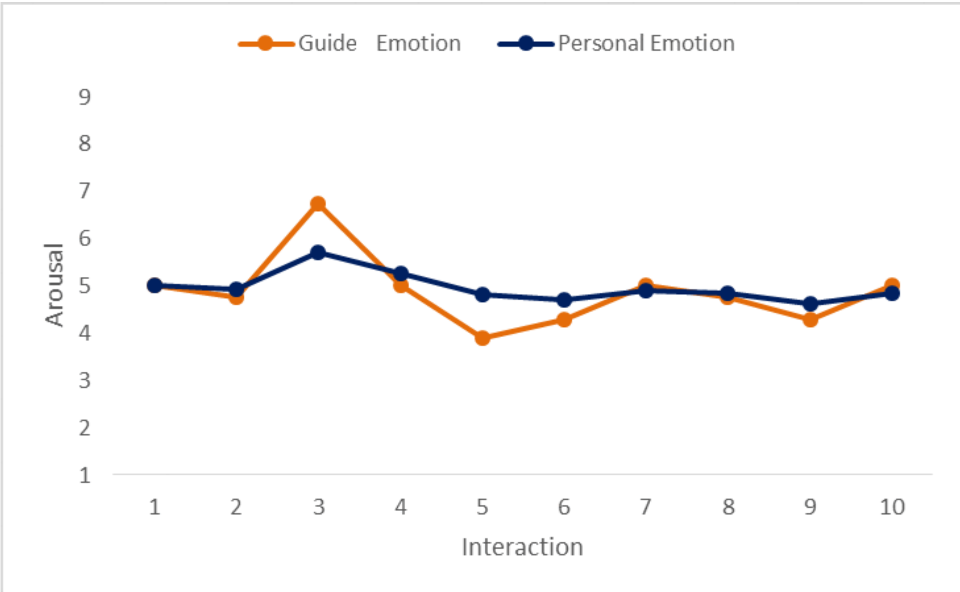
In this experiment, one robot is used for evaluating the modeling social effects in both perspective of sharing and referencing to robots. The robot personal emotion is set to neutral-low arousing whose values of valence and arousal are equal to 5 and 1 respectively. Human guided are provided directly during human-robot interactions. We selected 5 different images which consist one negative-low arousing image, one negative-high arousing image, two positive-low arousing images, and one neutral-mid arousing image, respectively. Those images are numbered from 1 to 5 and given to robot in a sequence of 5, 1, 2, 5, 3, 4, 5, 1, 4, and 5 for ten times. In this sequence, the robot sees images 5 for 4 times at interaction 1, 4, 7 and 10. Image 4 is seen in two times at interaction 6 and 9 while images 1 is seen at interaction 2 and 8. The robot sees images 2 and 3 for only one time at interaction 3 and 5 respectively.

5.3.2 Results

We set ϵ equal to 0.5 for applying Equation 4.5 to find the similar experiences of robots and acquiring new knowledge and interpretation about new stimuli from the environment during interactions with humans. For updating robot personal emotions in respond to those stimuli, we apply Equation 4.11, 4.12 and 4.13 through setting μ equal to 1 and 5 for valence and arousal respectively. The obtained result is shown in Figure 5.4. The orange line presents for guide emotions which the robot obtained from the human, and the dark blue line describes for personal emotions of robots. Both kinds of emotions are captured after every interaction between human and robot during 10 times of seeing visual stimuli.



(a) Valence component of guide emotions and personal emotions.



(b) Arousal component of guide emotions and personal emotions.

Figure 5.4: Guide emotions and personal emotions of the robot in the first experiment are captured after 10 interactions with humans to analyze the modeling of social effects in robot emotion generation. Each value is recorded after finishing one interaction which the robot sees a visual stimulus and gets human guide emotions for that stimulus. The robot sees image 5 in interactions 1, 4, 7 and 10. Images 1 and 4 are seen twice in interactions 6, 9 and 2, 8 respectively.

5.3.3 Discussion

In this experiment at the first interaction, the robot observes image 5 for the first time, it generates values of valence and arousal as similar as corresponding components of the guide emotion. This reaction is common in human infants as a result of social referencing when they watch their parents during interactions with them and then generate the same emotions through copying processes.

However, the generated emotions of robots in respond to images 1, 2, 3, and 4 when it sees them the first time at the interaction number 2, 3, 5 and 6 respectively are not the same as guide emotions. It can be explained with the change of the role of human guide to the robot emotion generation process. After the first interaction, the robot gains a certain knowledge about cats and form affective appraisal to the object by extracting visual features of the object based on human guide. Starting at the second interaction, the robot can use personal knowledge to negotiate with human guide for gaining a mode detailed interpretation about cat. At this stage, humans start to reduce the role of a teacher to increase a role of a partner to share knowledge with robot for influencing robot emotion generation direction. Thus, both role of human, teaching and sharing which are modeled based on the inspiration from human social effects, direct and shape the way robot interpret and generate emotions in respond to stimuli from the environment during human-robot interactions.

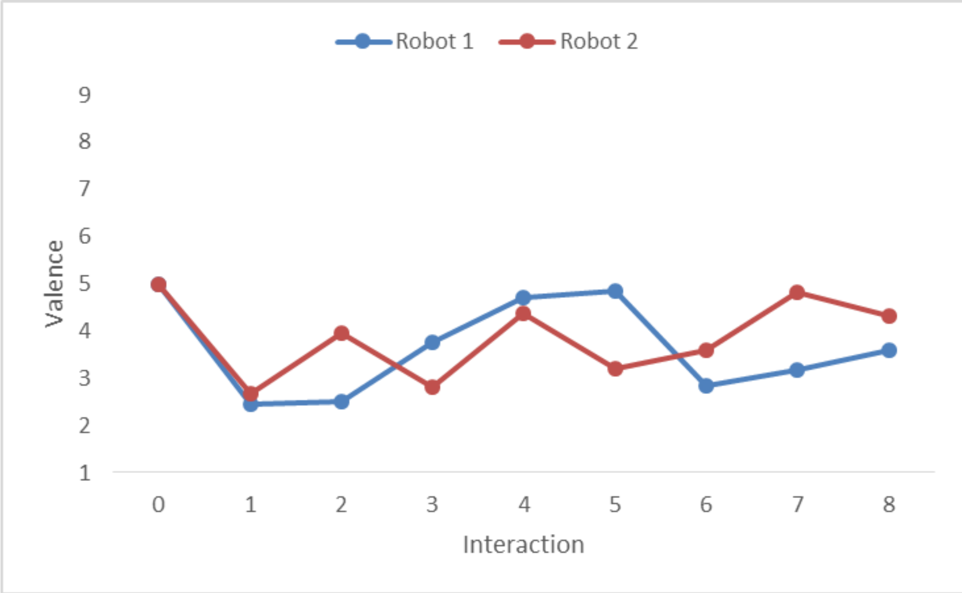
Furthermore, robot personal emotions in respond to image number 5 at interactions number 4, 7, and 10 are not the same as the guide emotion, which is different from the first time robot watch this image at interaction number 1. The emotional state of the robot in respond to image 5 during ten interactions are changed significantly from negative mid-arousing (2.94, 5.01) at the interaction number 4 to positive low-arousing (5.11, 4.85) at the interaction 10. This effect happens similarly when the robot sees images 1 twice at the interaction number 8 and image 4 twice at the interaction number 9. After seeing a certain stimulus at the first time, the robot can recognize the image in next interactions, so the robot can recall past knowledge about the image including experienced emotions. Thus, the robot is affected by the same experienced emotion and guide emotion in generating different personal emotion in respond to the stimulus because the emotional state of the robot at the time encoding the image are different.

5.4 Experiment 2: Robot Emotion Representation based on Multiple Cues

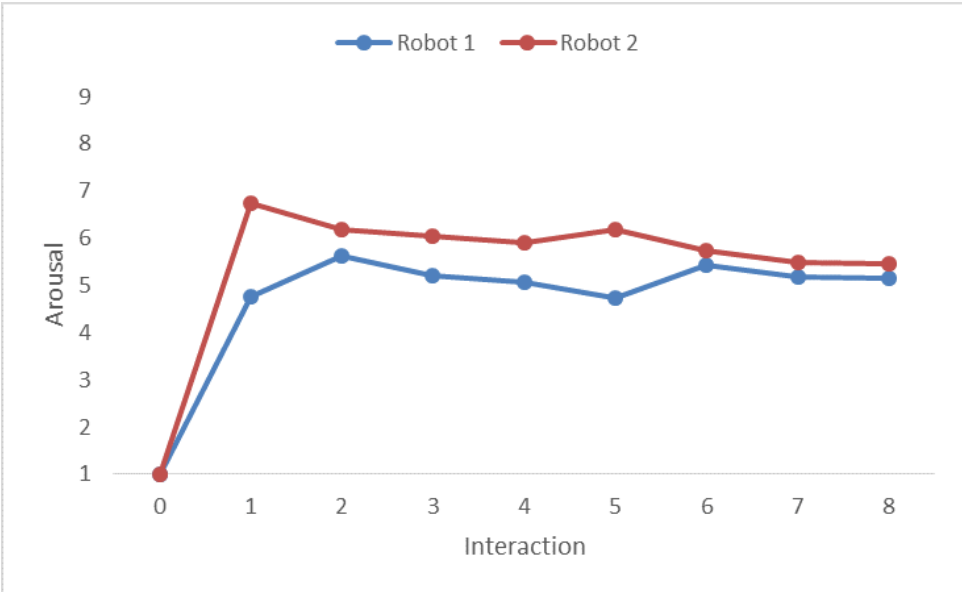
In this experiment, we expect our robots can generate emotions based on the influence from the integration of multiple cues. In that, robots may have different emotions in response to a certain stimulus at the same time of observation.

5.4.1 Design

Two robots are initialized with the same neutral-non arousing emotional state; in other words, initial values of valence and arousal of those robots are equal to 5 and 1 respectively. We provide two robots with the same guide from a human during the interactions with humans. We give 8 images numbered from 1 to 8 to our robots with different sequences. Those images contain one negative-low arousing image, one negative-high arousing image, four positive-low arousing images, and three neutral-mid arousing images, respectively. For the first robot, we show those images in a sequence of 1, 2, 3, 4, 5, 6, 7, and 8. The second robot observes images in the sequence of 2, 3, 1, 4, 6, 7, 5, and 8. In those sequences, two robots see image 4 and image 8 at the same time. Besides, before seeing image number 4, both robots are seen images 1, 2 and 3. Similarly, those robots see image number 8 after observing all other images.



(a) Valence component of Personal Emotions of two robots.



(b) Arousal component of Personal Emotions of two robots.

Figure 5.5: Personal Emotion of two robots in the second experiment captured for 8 interactions in which two robots see image 4 and 8 at the fourth and the eighth interactions respectively.

5.4.2 Results

The result of this experiment is shown in Figure 5.5. The blue line presents for the personal emotion of the first robot, and the red line describes the personal emotion of the second robot We apply the same values of ϵ and μ in the first experiment for Equation 4.11, 4.12 and 4.13 in order to form object appraisals for evaluating similarity between seen objects and new objects and updating robot personal emotions. The result shows the change of personal emotions of both robots during eight times of observing visual images with human assist. We record robot personal emotions after being affected by a certain visual stimulus in one interaction with the same human. At interaction 0, both robots are at the initial states which equal to neutral-mid arousing emotion. The value of valence and arousal are 5 and 1 respectively.

5.4.3 Discussion

The generated emotions of two robots in respond to image 4 and 8 for the first time of being affected at the same interaction are different even though they are initiated by the same emotional state and provided by the same guide emotions from the human. On the one hand, the first robot responds to image 4 and 8 via negative-mid arousing emotion and high negative-mid arousing emotion respectively. On the other hand, the second robot generates a lower values of valence and a higher values of arousal for image 4 while responding an emotion which is less negative and higher arousing for image 8 than the generate emotion of the first robot.

The reason is because images are given to both robots are the same but in different sequences; thus, at a certain time, the gained knowledge of each robot is different from the other. For example, the first robot gained knowledge from images 1 and 2 while the second robot achieved experiences from images 2 and 3 after the second interaction. The different obtained knowledge leads to forming different object appraisals for the same category of image, cat, and generate different emotions in respond to those objects. After the third interactions, both robots are seen image 1, 2, and 3. However, their emotions in respond to those images are different. For instance, the first robot responds high negative-mid arousing to image 1 at interaction 1 while the second robot generate negative-high arousing emotion to the same image at the interaction 3. At the fourth interaction, both robots saw image 1, 2 and 3 with different sequences and gained different experiences including engaged emotions. Similarly for image 8, both robots observed images 1, 2, 3, 4, 5, 6, and 7 in previous interactions but in different sequences, thereby forming different personal experiences. When the retrieval process occurs, each robot recall different experienced emotions for the same object which is mostly similar to the new object to generate responded emotions.

In additions, personal emotions of two robots at the time of encoding the images 4 and 8 are different from each other, which also influence to the generation of emotions of two robots for those images. Different past experiences and different personal emotions at the time encoding a new experience enable robots to have different emotions in respond to the same stimulus, which may help to the formation of the robot personality.

5.5 Conclusion

Our two experiments are designed based on the proposed scenario which is used to evaluate how our system works and the model of our robot emotion representation based on human psychology theories. Those experiments show that the proposed emotion representation provides robots with a capability to generate emotions similar to the way humans do based on social effects and past experiences. Robots play two roles of a passive learner and an active partner to be influenced by human-inspired social effect by both perspective of referencing and sharing. Moreover, robots are able to form affective appraisals for gaining knowledge and recalling past experience in acquiring a more detailed interpretation about the environment with social context which is shaped and direct from humans; thereby generating emotions due to human direction. The second experiment strongly presents that different prior interpretation about the same stimuli can lead to form different knowledge and generate emotions even the resource of information are unchanged. Two different robots generate different emotions for the same stimulus due to their own personal knowledge which is taught and shared by humans equally.

Chapter 6

Summary

6.1 Conclusion

The limitation of modeling emotions for robots is addressed with the review on previous works presented in Chapter 2. Artificial emotions are modeled to robots categorized into basic emotions and multiple dimensional representation of emotions. In addition, robots cannot acquire knowledge continuously due to the change of the environment because of limited functionality, some cognition simulation, the use of limited basic emotions on directing behaviors and expressions; thereby the interactions between humans and robots is not effective after a long-term of engagement. We investigate emotion activation cues during human-human interactions which are social effects and past experiences presented in Chapter 3 to propose our approach on representing robot emotions which would be more likely accepted by humans. Therefor, enhancing a believable interaction between human and robot which can stand for a long-term of engagement.

In Chapter 4, a robot long-term memory architecture is designed based on the adoption from a robot developmental memory architecture for enabling our robots to acquire knowledge intelligently in general instead of simulating some specific case of human cognitive processes. We also describe the approach of using a recent Machine Learning technique, Convolutional Neural Network to model human-inspired social effects in our robot emotion representation. However, the influence of social effects based on human culture cannot be modeled due to the recent approach, thus we allow our robots to get human guide directly through HRI. In this chapter, the integration of human-inspired multiple emotion activation cues is presented for providing robots with a capacity to acquire knowledge continuously, generate emotions and represent those emotions on the valence-arousal coordinate during interactions with humans and the environment .

We present our designed scenario and two main experiments in order to investigate the performance of our approach on representing robot emotions in Chapter 5. The first experiment shows that our approach can model both perspectives of referencing and sharing of human-inspired social effects in generating robot emotions. Robots can play a role as an passive learner to accept human guide and learn from that to acquire knowledge. The obtained experiences are used for negotiating with human guide for acquiring a more detailed personal interpretation of robots about stimuli in the environment during HRI. For a certain stimulus, one robot can generate different emotions due to the influence of past experiences and emotional state at the time of encoding the experience when getting the same guide from a human. The second experiment describes that our robots are able to form personality on generating emotions and acquiring knowledge. In other words, two different robots provided by the same human guide can generate different emotions for the same stimulus at the same time of observing.

6.2 Contribution

We designed our robot long-term memory based on the adoption from a developmental memory architecture through epigenetic approach to solve problem of simulating human specific cognition. Besides, we also consider the enhancement role of emotions in memory and the effect of past experiences to emotion generation through memory retrieval for modeling our robot long-term memory. The memory architecture allows robots to form affective appraisals for acquiring the interpretation about a novel object or stimulus in order to gain experiences continuously, generate emotions and adapt intelligently to stimuli from the environment during interactions with humans.

In attempt of modeling human-inspired social effects, we point out that our current approach on using Machine Learning techniques to provide robots with a capacity to recognize humans emotion from natural images is not suitable when ignoring the role of human culture. Thus, we model the influence of social effects to our robots by giving robots human guide emotion directly through HRI. Both aspects of social effects are fully considered due to referencing and learning; therefore, our model enable robots to form personal knowledge, emotions and personality. Robots play not only a role as a passive learner to accept guide from humans but also a role as an active partner to be directed by human guide. Robot personal knowledge and emotions are shaped and influenced by humans, which would be accepted by humans positively.

An experiment scenario is presented to design two main experiments for evaluating our robot emotion representation performance. The modeling of social effects in both perspectives of referencing and sharing are considered in the first experiment by using one robot. Several interactions are made between a human and the robot, emotions of the robot are recorded after each interaction in order to evaluate. Psychology experiments are used to analyze the results, thereby showing that robots learn from humans due to social effects similarly in infants for a very initial state and likely older children after gaining personal knowledge. The second experiment aims at evaluating the integration of multiple cues in robot emotion generation by using two robots. Similar to the first experiment, robot emotions are obtained after each interactions. The result shows that obtained personal knowledge can influence strongly in robot emotion generation and knowledge acquirement which can be considered as an important factor in forming personality.

6.3 Future work

The current model is limited at detecting one type of object, human emotion recognition and object localization. Besides, there is no related works which can be used for evaluating our current results at the time this research has been finished . In the future we propose several approaches to improve and evaluate our model presented as following.

- Implementing the proposed model on real robots and exploring the natural interaction between humans and robots. Based on experiments on a real robots with humans, we can evaluate how well this model work in a real life and improve the emotion representation based on the feedback of participants. In addition, robots should interpret human facial expression or gestures to get the emotional guide message as seen in human infants to generate emotions and obtain personal knowledge.

- In order to detect and recognize more types of objects, robots should be provided with a powerful function to extract more features of the objects. The current function enable robots to get color and shape features of objects. Robots are not able to distinguish the different between dog and cat.

- The robot is dependent on humans through interactions to gain knowledge and generate emotions. Machine Learning technique can be used to model the role of human guide which can investigate the role of human culture in generating emotions; thereby enabling robots to acquire knowledge and experiences independently.
- An evaluation matrix is necessary to design for evaluating the influence of each human-inspired cue which activates robot emotion. Due to this matrix, we can rank those cues due to their influence.
- There are many external factors from the environment and internal factors inside humans such as personality or mood can be considered in the future for modeling robot emotions.
- Other kind of input data can be investigated such as sensor information which helps robots to interpret the heat temperature of humans for recognizing human guide emotions. Besides, speech and natural language processing should be considered to allow robots to understand human guide naturally.

Bibliography

[1] https://en.wikipedia.org/wiki/Robert_Plutchik.

[2] http://www.cs.ru.ac.za/research/g02m1682/.

[3] James S Adelman and Zachary Estes. Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, 129(3):530–535, 2013.

[4] Hyeon Min Ahn, Shin Ah Kim, In Jae Hwang, Ji Woon Jeong, Hyun Taek Kim, Stephan Hamann, and Sang Hee Kim. The effect of cognitive reappraisal on long-term emotional experience and emotional memory. *Journal of neuropsychology*, 9(1):64–76, 2015.

[5] Christopher Allan, Ruby Valverde Ibánez, Matthias U Keysermann, and Patrıcia A Vargas. *Implementing a data mining approach to emotional memory modelling for robocup players*. PhD thesis, MSc Thesis, 2013.

[6] Albert Bandura. Social cognitive theory of social referencing. In *Social referencing and the social construction of reality in infancy*, pages 175–208. Springer, 1992.

[7] Lisa Feldman Barrett, Batja Mesquita, Kevin N Ochsner, and James J Gross. The experience of emotion. *Annual review of psychology*, 58:373, 2007.

[8] Patricia J Bauer and Louise A Hertsgaard. Increasing steps in recall of events: Factors facilitating immediate and long-term memory in 13.5-and 16.5-month-old children. *Child development*, 64(4):1204–1223, 1993.

[9] Paul Baxter, Tony Belpaeme, Lola Canamero, Piero Cosi, Yiannis Demiris, Valentin Enescu, A Hiolle, I Kruijff-Korbayova, R Looije, M Nalin, et al. Long-term human-robot interaction with young users. In *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop)*, 2011.

[10] Francisco Bellas, Andrés Faiña, Gervasio Varela, and Richard J Duro. A cognitive developmental robotics architecture for lifelong learning by evolution in real robots. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2010.

[11] Jonah Berger. Arousal increases social transmission of information. *Psychological science*, 22(7):891–893, 2011.

[12] Marian E Berryhill and Ingrid R Olson. Is the posterior parietal lobe involved in working memory retrieval?: Evidence from patients with bilateral parietal lobe damage. *Neuropsychologia*, 46(7):1775–1786, 2008.

[13] Isabelle Blanchette. *Emotion and reasoning*. Psychology Press, 2013.

[14] Jeffrey A Brooks, Holly Shablack, Maria Gendron, Ajay B Satpute, Michael H Parrish, and Kristen A Lindquist. The role of language in the experience and perception of emotion: a neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience*, page nsw121, 2016.

[15] Michel Cabanac. What is emotion? *Behavioural processes*, 60(2):69–83, 2002.

[16] Martin A Conway. Sensory–perceptual episodic memory and its context: Autobiographical memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1413):1375–1384, 2001.

[17] Antonio Damasio. Descartes's error: emotion. *Reason and the Human Brain, Grosset and Putnam, New York*, 1994.

[18] Elisa Calhau De Castro and Ricardo Ribeiro Gudwin. An episodic memory for a simulated autonomous robot. *Proceedings of robocontrol*, pages 1–7, 2010.

[19] Lycia D de Voogd, Guillén Fernández, and Erno J Hermans. Disentangling the roles of arousal and amygdala activation in emotional declarative memory. *Social cognitive and affective neuroscience*, page nsw055, 2016.

[20] Will Dodd and Ridelto Gutierrez. The role of episodic memory and emotion in a cognitive robot. In *Proceedings of 14th Annual IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, pages 692–697, 2005.

[21] Paul Ekman, Wallace V Friesen, Maureen O'Sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E Ricci-Bitti, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712, 1987.

[22] Saul Feinman. Social referencing in infancy. *Merrill-Palmer Quarterly (1982-)*, pages 445–470, 1982.

[23] Saul Feinman, Debra Roberts, Kuei-Fang Hsieh, Darby Sawyer, and Dee Swanson. A critical review of social referencing in infancy. In *Social referencing and the social construction of reality in infancy*, pages 15–54. Springer, 1992.

[24] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166, 2003.

[25] Jaclyn Hennessey Ford, John A Morris, and Elizabeth A Kensinger. Effects of emotion and emotional valence on the neural correlates of episodic memory search and elaboration. *Journal of cognitive neuroscience*, 26(4):825–839, 2014.

[26] Steven Handel. Classification of emotions, 2012.

[27] Paul L Harris, Carl N Johnson, Deborah Hutton, Giles Andrews, and Tim Cooke. Young children's theory of mind and emotion. *Cognition & Emotion*, 3(4):379–400, 1989.

[28] Russell JA. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.

[29] Eun-Sook Jee, Chong Hui Kim, Soon-Young Park, and Kyung-Won Lee. Composition of musical sound expressing an emotion of robot based on musical factors. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 637–641. IEEE, 2007.

[30] Zerrin Kasap and Nadia Magnenat-Thalmann. Towards episodic memory-based long-term affective interaction with a human-like robot. In *19th International Symposium in Robot and Human Interactive Communication*, pages 452–457. IEEE, 2010.

[31] Kazuhiko Kawamura. Cognitive approach to a human adaptive robot development. In *Proceedings of 14th Annual IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005)*, pages 629–634, 2005.

[32] Tatsuhiro Kishi, Nobutsuna Endo, Takashi Nozawa, Takuya Otani, Sarah Cosentino, Massimiliano Zecca, Kenji Hashimoto, and Atsuo Takanishi. Bipedal humanoid robot that makes humans laugh with use of the method of comedy and affects their psychological state actively. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1965–1970. IEEE, 2014.

[33] Tatsuhiro Kishi, Takuya Otani, Nobutsuna Endo, Przemyslaw Kryczka, Kenji Hashimoto, K Nakata, and Atsuo Takanishi. Development of expressive robotic head for bipedal humanoid robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4584–4589. IEEE, 2012.

[34] Mary D Klinnert. The regulation of infant behavior by maternal facial expression. *Infant Behavior and Development*, 7(4):447–465, 1984.

[35] Mary D Klinnert, Robert N Emde, Perry Butterfield, and Joseph J Campos. Social referencing: The infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology*, 22(4):427, 1986.

[36] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[37] Kevin S LaBar and Elizabeth A Phelps. Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, 9(6):490–493, 1998.

[38] Kristin Hansen Lagattuta and Henry M Wellman. Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion. *Child Development*, 72(1):82–102, 2001.

[39] Kristin Hansen Lagattuta, Henry M Wellman, and John H Flavell. Preschoolers' understanding of the link between thinking and feeling: Cognitive cuing and emotional change. *Child development*, 68(6):1081–1104, 1997.

[40] Barbara Landau, Linda Smith, and Susan Jones. Object shape, object function, and object name. *Journal of Memory and Language*, 38(1):1–27, 1998.

[41] Richard D Lane, Lee Ryan, Lynn Nadel, and Leslie Greenberg. Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, 38:e1, 2015.

[42] Peter J Lang, Margaret M Bradley, and Bruce N Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. *Technical report A-8*, 2008.

[43] Richard S Lazarus. Progress on a cognitive-motivational-relational theory of emotion. *American psychologist*, 46(8):819, 1991.

[44] Francis Leconte, François Ferland, and François Michaud. Design and integration of a spatio-temporal memory with emotional influences to categorize and recall the experiences of an autonomous mobile robot. *Autonomous Robots*, 40(5):831–848, 2016.

[45] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[46] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[47] Xin Lu, Poonam Suryanarayan, Reginald B Adams Jr, Jia Li, Michelle G Newman, and James Z Wang. On shape and the computability of emotions. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 229–238. ACM, 2012.

[48] Artur Marchewka, Łukasz Żurawski, Katarzyna Jednoróg, and Anna Grabowska. The nencki affective picture system (naps): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior research methods*, 46(2):596–610, 2014.

[49] Raja Majid Mehmood and Hyo Jong Lee. Erp analysis of emotional stimuli from brain eeg signal. In *Int'l Conf. Biomedical Engineering and Science*, page 5, 2015.

[50] Anthony F Morse, Joachim De Greeff, Tony Belpeame, and Angelo Cangelosi. Epigenetic robotics architecture (era). *IEEE Transactions on Autonomous Mental Development*, 2(4):325–339, 2010.

[51] Brendan D Murray, Alisha C Holland, and Elizabeth A Kensinger. Episodic memory and emotion. *Handbook of cognition and emotion*, pages 156–175, 2013.

[52] Lynn Nadel, A Samsonovich, Lee Ryan, and Morris Moscovitch. Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*, 10(4):352–368, 2000.

[53] Henrik Nordström and Stefan Wiens. Emotional event-related potentials are larger to figures than scenes but are similarly reduced by inattention. *BMC neuroscience*, 13(1):1, 2012.

[54] Hadas Okon-Singer, Jan Mehnert, Jana Hoyer, Lydia Hellrung, Herma Lina Schaare, Juergen Dukart, and Arno Villringer. Neural control of vascular reactions: impact of emotion and attention. *The Journal of Neuroscience*, 34(12):4251–4259, 2014.

[55] Andrew Ortony, Gerald L Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.

[56] W Gerrod Parrott. *Emotions in social psychology: Essential readings*. Psychology Press, 2001.

[57] James W Pennebaker, Emmanuelle Zech, Bernard Rimé, et al. Disclosing and sharing emotion: Psychological, social, and health consequences. *Handbook of bereavement research: Consequences, coping, and care*, pages 517–543, 2001.

[58] Wendy Phillips, Simon Baron-Cohen, and Michael Rutter. The role of eye contact in goal detection: Evidence from normal infants and children with autism or mental handicap. *Development and Psychopathology*, 4(03):375–383, 1992.

[59] Joanna Pilarczyk and Michał Kuniecki. Emotional content of an image attracts attention more than visually salient features in various signal-to-noise ratio conditions. *Journal of vision*, 14(12):4–4, 2014.

[60] Robert Plutchik and Henry Kellerman. *Theories of emotion*, volume 1. Academic Press, 2013.

[61] Ferdian Pratama, Fulvio Mastrogiovanni, Soon Geul Lee, and Nak Young Chong. Long-term knowledge acquisition using contextual information in a memory-inspired robot architecture. *Journal of Experimental & Theoretical Artificial Intelligence*, pages 1–22, 2016.

[62] Kanlaya Rattanyu and Makoto Mizukawa. Emotion recognition based on ecg signals for service robots in the intelligent space during daily life. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 15(5):582–591, 2011.

[63] Nathaniel Rochester, J Holland, L Haibt, and W Duda. Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on information Theory*, 2(3):80–93, 1956.

[64] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

[65] James A Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.

[66] James A Russell and Merry Bullock. Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults. *Journal of Personality and Social Psychology*, 48(5):1290, 1985.

[67] Jelle Saldien, Kristof Goris, Bram Vanderborght, Johan Vanderfaeillie, and Dirk Lefeber. Expressing emotions with the social robot probo. *International Journal of Social Robotics*, 2(4):377–389, 2010.

[68] Gregory R Samanez-Larkin, Elaine R Robertson, Joseph A Mikels, Laura L Carstensen, and Ian H Gotlib. Selective attention to emotion in the aging brain. 2014.

[69] T. Shibata. Mental commit robot (paro). `http://www.paro.jp/`.

[70] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[71] Larry R Squire and Pablo Alvarez. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current opinion in neurobiology*, 5(2):169–177, 1995.

[72] Alexander Stoytchev. Some basic principles of developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 1(2):122–130, 2009.

[73] Jennifer M Talarico, Kevin S LaBar, and David C Rubin. Emotional intensity predicts autobiographical memory experience. *Memory & cognition*, 32(7):1118–1132, 2004.

[74] Andrea Lockerd Thomaz, Matt Berlin, and Cynthia Breazeal. An embodied computational model of social referencing. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 591–598. IEEE, 2005.

[75] Kazuyoshi Wada and Takanori Shibata. Social and physiological influences of living with seal robots in an elderly care house for two months. *Gerontechnology*, 7(2):235, 2008.

[76] Tedra A Walden and Tamra A Ogan. The development of social referencing. *Child development*, pages 1230–1240, 1988.

[77] Arlene S Walker-Andrews. Infants' perception of expressive behaviors: differentiation of multimodal information. *Psychological bulletin*, 121(3):437, 1997.

[78] Juyang Weng. Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, 1(02):199–236, 2004.

[79] Paul Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. 1974.

Publications

International Conferences

[1] Dang, T.L.Q., Jeong, S. and Chong, NY. "Personalized Robot Emotion Representation through Retrieval of Memories" *Proceedings of International Conference on Control, Automation and Robotics (ICCAR)*, IEEE , Nagoya, Japan, April, 2017.