A Framework for Recognizing and Regulating Emotions in the Elderly

José Carlos Castillo[1], Antonio Fernández-Caballero[2], Álvaro Castro-González[1], Miguel A. Salichs[1], and María T. López[2]

[1] Universidad Carlos III de Madrid, Robotics Lab, 28911-Madrid, Spain
{jocastil,acgonzal,salichs}@ing.uc3m.es
[2] Universidad de Castilla-La Mancha, Instituto de Investigación en Informática de Albacete, 02071-Albacete, Spain
{Antonio.Fdez,Maria.LBonal}@uclm.es

Abstract. This paper introduces a gerontechnological framework which enables real-time and continuous monitoring of the elderly and provides the best-tailored reactions of a social robot and the proper ambience in order to regulate the older person's emotions towards a positive emotion. After describing the benefits of the framework for emotion recognition and regulation in the elderly, the eight levels that compose the framework are described. The framework recognizes emotions through studying physiological signals, facial expression and voice. Emotion regulation is enabled by tuning music, color and light to the specific need of the elderly.

Keywords: Gerontechnological Framework, Emotion Regulation, Emotion Recognition, Elderly Monitoring, Social robotics.

1 Introduction

Developed countries are dealing with the effects of population ageing, generally due to lower birth rates and higher life expectancy. Population ageing is defined as a shift in the distribution of a country's population towards greater ages [1]. Elderly frequently have special needs and require a close and personalized monitoring, specifically at home and mainly due to health-related issues [2], [3]. Indeed, there is a growing trend towards ambient assisted living approaches which provide support to the older people in a personalized way. Also, the inclusion of companion robots in the intelligent habitat of the older person is a solution towards quality of life of the elderly living alone [4], [5]. In recent years, the robotics community has seen a gradual increase in social robots, that is, robots that exist primarily to interact with people [6]. We believe that perceiving / enhancing the quality of life of the elderly living at home is possible through automatic emotion recognition and regulation (from now on ERR). Indeed, it has been largely studied that positive emotional states promote healthy perceptions, beliefs, and physical well-being [7].

L. Pecchia et al. (Eds.): IWAAL 2014, LNCS 8868, pp. 320–327, 2014.
© Springer International Publishing Switzerland 2014

Existing emotion recognition technologies are divided into three major categories depending on what kind of data is analyzed: physiological signals, facial expressions, or voice [8]. Physiological emotion recognition shows acceptable performance but has some critical weaknesses that prevent its widespread use; they are obtrusive to users and need special equipment or devices [9]. Facial analysis includes a number of processing steps which attempt to detect or track the face, to locate characteristic facial regions such as eyes, mouth and nose on it, to extract and follow the movement of facial features. Most of the proposed works are based on the Facial Action Coding System (FACS), based on the definition of "action units" of a face that cause facial movements [10]. Lastly, the speech signal conveys a large amount of information. Two main problems are addressed: (1) finding the set of features in the speech signal that are most significant in conveying emotions and (2) finding the best classification algorithm that can indicate emotional expression, based on the above features [11].

Now, emotion regulation refers to a set of processes that either stop the emotion from emerging or prevent it from being expressed once it is triggered [12]. Bottom-up emotion generation refers to the elicitation of emotion by the presentation of a stimulus that is thought to have simple physical properties that are inherently emotional [13]. Antecedent emotion regulation strategies apply while the emotion is still unfolding and has not reached its peak. In our case, we are interested in response-focused emotion regulation, which tries to aim at altering and controlling the experiential, behavioral and physiological response to the fully established emotion. There is a great deal of work based on the use of music [14], color [15] and lighting [16] in emotion regulation.

This paper discusses the need of frameworks capable of accepting any sensor and actuator dealing with the problem of detecting and regulating emotions through ambient intelligence and social robotics. In this sense, we introduce a gerontechnological framework that enables continuously monitoring the elderly and provides the best-tailored reactions of a social robot and the proper ambience in order to regulate their emotions towards a positive emotional state.

2 Why a Framework?

Nowadays, frameworks are playing a critical role in complex software systems development. A software framework is a set of software blocks that programmers can use, extend or modify to fit specific applications. This mechanism allows the decomposition of an application into a set of independent modules describing a set of interfaces to ease communications. Therefore, frameworks constitute a flexible approach that enables the quick creation and deployment of applications. Under this scheme, components are defined as reusable and independent blocks that are combined to other components in order to build a specific system. From a developer perspective, building an application consists of assembling some existing components and a few of their own.

Among the number of advantages, the biggest one is that a software framework reduces the time in developing any software as it provides a set of blocks users can directly use or improve. This means that users can be abstracted from the lower level implementation and focus the effort on programming their required module. Another important feature provided by frameworks is the use of design patterns, which makes code cleaner and more extensible for future requirements.

The main disadvantage of frameworks is the development complexity. For small projects, it is faster to code directly rather than implement a whole framework infrastructure. In contrast, for big projects, like the one proposed in this paper, a framework is an appropriated option.

Real time ERR involves a wide diversity of technologies. This will result in a complex system which will require a powerful, flexible framework. Designing a new framework presents crucial challenges:

Deciding the purpose of the framework. This simple decision involves several issues related to the scope of the problem to be solved. For instance, a framework for web development has nothing to do with the one presented in this paper, so the requirements of each must be carefully analyzed before taking further steps.

Assessing the levels. After having confirmed the purpose of the framework, it is necessary to establish a division to allow a structured development. Finding an optimal decomposition involves both a deep analysis of the problem as well as subject matter experts.

Determining the grain size. This problem is related to the levels one. At an early design stage, the functionalities included in each level have to be clearly defined at design time.

Avoiding being restrictive. Finally, a framework must be adaptable to a growing problem, but not too wide to loose its purpose. This is an important trade-off that must also be considered at the design stage.

3 Other Frameworks in ERR

Frameworks are widely spread for web-based environments or user-oriented applications. Some examples are *Grails* or *Springs* to develop Java-based applications, or *ASP.NET* for Windows-based ones. The academic community has adopted such paradigm for research projects related to ERR.

Affect sensing by machines has been argued as an essential part of next-generation human-computer interaction (HCI). To this end, in the recent years a large number of studies have been conducted, which report automatic recognition of emotion as a difficult, but feasible task. However most effort has been put towards off-line analysis, whereas to date only few applications exist able to react to a user's emotion in real-time. The framework called Smart Sensor Integration considerably boosted the development of multimodal on-line emotion recognition systems [17]. Furthermore, Gehrig and Ekenel presented a common framework for real-time action unit detection and emotion recognition [18]. For these tasks, they employed a local appearance-based face representation approach using discrete cosine transform. Santos et al. came up with a proposal that performed

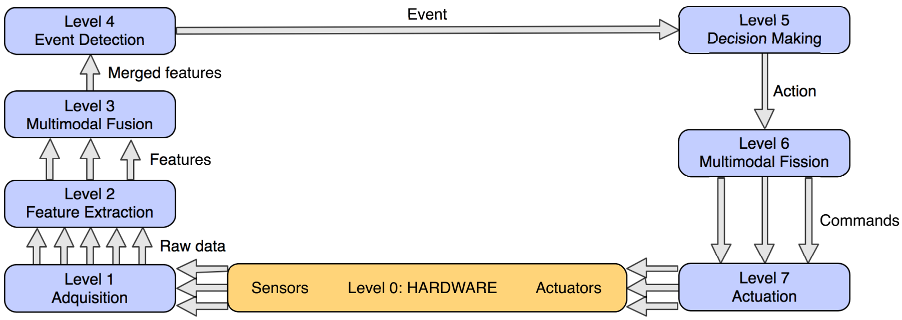

Fig. 1. The levels of the proposed framework

real-time stress detection based only in two physiological signals: heart rate and galvanic skin response [19].

In relation to the emotion regulation, the literature does not offer many software frameworks. The emotion regulation will be inspired by works from the area of psychology. Psychologists have explored traditional frameworks (defined as a set of concepts, practices, and criteria to deal with a particular kind of problem) for diagnosis and treatment of deceases. An example is the work presented by Werner and Gross provides a framework for researchers and clinicians interested in understanding the role of emotion regulation processes in psychopathology [20]. We will use these works to provide the proper reaction of the social robot and the ambience.

4 A New Framework for ERR

The proposed framework is intended to deal with the problems that can be found in real environments (hardware abstraction, noise in the data, or real-time processing). In this section we provide a high level description of the framework and the functions performed in each level. An overview of the framework we propose is shown in Fig. 1.

Each level is defined by the information that it consumes (receives), how this incoming data is transformed, and the resulting data it provides (sends) to the next level. Therefore these data flows are detailed for each level in the next subsections. Fig. 2 shows the outstanding elements to be considered in the proposed framework for ERR.

Level 0: Hardware - Sensors and Actuators. Frameworks are formed by software elements located at different levels. However, when dealing with data coming from the real world, also referred as the environment, we need to place the hardware elements at some level. In our case, this is Level 0. This layer represents the hardware components in charge of perceiving the environment

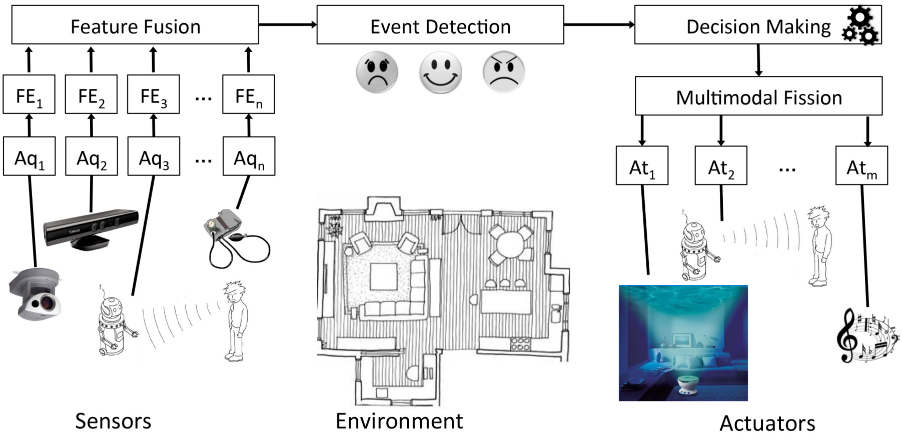

Fig. 2. Outstanding elements that are described in the proposed framework

and performing actions. No software module is considered here although the software from other levels directly interacts with it.

Sensors are devices that sense the environment and convert the information into electronic format. They can be roughly classified according to the information they provide, among others, visual data (RGB cameras, IR cameras, or depth sensors), audio (microphones), biometric data (blood pressure, perspiration, breathing rate, and so on). Actuators are devices able to alter the environment by receiving instructions and transforming them into actions. Some actuators considered in this project can be: projectors, lights, a robot, a sound system, or a heating system.

Level 1: Acquisition. Software at this level handles the communication with the sensors. For each sensor, there is a module in this level that understands the data received. Since this is the lowest processing level, it does not consume data from other levels. It reads the low-level, raw data from the sensors and provides it the next levels.

Level 2: Feature Extraction. This is one of the key levels. Here the raw data provided by the sensors is interpreted by different algorithms. This means that the input data is transformed into more informative, numerical features that summarize the raw data. Each feature represents an important characteristic of the data that is of special interest.

From a computer vision perspective, this level is in charge of detecting, recognizing and tracking those objects of interest contained in the scene. In this sense, features may be understood as the track and shape of the object, as well as the kind of object detected. Other example, some features coming from the elder's utterances can be extracted, such as the sentiment (positive, negative, neutral), the gender of the speaker, or the topics involved within the speech. So, this level consumes raw data and provides features to the upper level that will have to interpret them.

Level 3: Multimodal Fusion. As the system will not be able to completely rely on the separate sources (noise, false positives, or occlusions), a fusion step is crucial in order to reduce the uncertainty associated to the extracted features. Indeed, feature extraction can be performed over data coming from very different sources. In our scenario, we have information coming from cameras, wearable sensors sending biometric data, microphones, and others. For each kind of data, different features are extracted. All these heterogeneous features are combined in order to provide higher level abstraction of information. This is known as the multimodal fusion.

In our particular problem, the recognition of the elder's emotion in the scenario is crucial. For this task, our framework uses features related to the user's speech, their blood pressure, and their face. The fusion of all these features results on richer information at the time of assessing the elder's emotion. This level consumes the features extracted in Level 2, and provides a combination of feature in a time window. It is important to consider the time constrains in order to determine how often the fusion process is performed. For example, if it is detected that the elder is talking about a sad event, how long do we have to consider this feature? Therefore, the frequency of the fusion process will determine the lifespan of the features. The fusion process gives temporal coherence to the features.

Level 4: Event Detection. This level receives a set of features and processes them. The result is a high-level description of the situation (e.g. the elder is happy). In our problem, the emotion analysis logic is in this level. Its inputs are the array of features sent by the multimodal fusion. The output is the events related to the emotion of the elder. This means that the emotion recognition algorithm corresponds here. Moreover, our framework is not limited to detect user's emotion. The algorithms intended to detect other events have to be implemented at this level and the inputs would be the same.

Level 5: Decision Making. Here is where the *intelligence* of the system lies: considering the events published by the Level 4, the system reacts and decides the action to execute. For example, considering that our system reacts to the elder's mood, once this is detected, the system can vary the illumination, move the shutters, change the music, as well as the robot can interact with them in a proper way. This level consumes events and provides actions to be executed by the different actuators in the scenario.

Level 6: Multimodal Fission. This level defines the particular result of an action. A multimodal fission component accepts actions and split it into several commands. Each command is directed to a different actuator in the environment. For instance, if the action is to create a relaxing environment, the multimodal fission will publish commands addressed to dim the lights, project warm colors, play pleasant music, and the robot will have a conversation about the weather.

The fission process is tuned according to the user's likes. Hence, some information associated to the user is welcomed. Knowing some of their preferences some environmental conditions, such as the background music and the projections, can be selected accordingly. This results in a faster achievement of the desired effect. This level consumes actions from the decision making level, and provides commands to the proper actuators.

Level 7: Actuation. This level is the counterpart of Level 1. Here the modules communicating with the actuators send the proper commands. There is one module in charge of managing each actuator. The modules at this level consume commands published by the Fission level. Meanwhile, there is not information sent to other levels but they write low level instructions, or primitives, to hardware components. As an example, the text-to-speech module receives the command *say hello* that is transformed into low level instructions that synthesize "hello" through the speakers.

5 Conclusions

This paper has justified the need of a specific framework for emotion recognition and regulation tasks. Beside, we have introduced a new gerontechnological framework for monitoring the elderly at home. In first place, the framework is aimed at detecting the elder's emotions by analyzing their physiological signals, facial expression and voice. Then, the framework provides the best-tailored reactions of a social robot and the ambience to regulate the elder's emotions towards a positive emotion. The current state of the art in emotion regulation through music, color and light is used by the framework with the final goal of enhancing the quality of life of elder people living alone at their homes. After describing the benefits of a framework for ERR in the elderly, the eight levels that compose the proposed framework have been described.

Acknowledgements. This work was partially supported by Spanish Ministerio de Economía y Competitividad / FEDER under TIN2013-47074-C2-1-R and TIN2010-20845-C03-01 grants. José Carlos Castillo was partially supported by a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism, operated by Universidad Complutense de Madrid.

References

1. Castillo, J.C., Carneiro, D., Serrano-Cuerda, J., Novais, P., Fernández-Caballero, A., Neves, J.: A multi-modal approach for activity classification and fall detection. International Journal of Systems Science 45(4), 810–824 (2014)
2. Costa, Â., Castillo, J.C., Novais, P., Fernández-Caballero, A., Simoes, R.: Sensor-driven agenda for intelligent home care of the elderly. Expert Systems with Applications 39(15), 12192–12204 (2012)

3. Dosi, G.: Sources, procedures and microeconomics effects of innovation. Economic Literature 26, 1120–1171 (1998)
4. Gascueña, J.M., Garijo, F.J., Fernández-Caballero, A., Gleizes, M.P., Machonin, A.: Deliberative control components for eldercare robot team cooperation. Journal of Intelligent and Fuzzy Systems (2014), doi:10.3233/IFS-141199
5. Martínez-Gómez, J., Fernández-Caballero, A., García-Varea, I., Rodríguez, L., Romero-González, C.: A taxonomy of vision systems for ground mobile robots. International Journal of Advanced Robotic Systems 11, 111 (2014), doi:10.5772/58900
6. Kirby, R., Forlizzi, J., Simmons, R.: Affective social robots. Robotics and Autonomous Systems 58, 322–332 (2010)
7. Salovey, P., Rothman, A.J., Detweiler, J.B., Steward, W.T.: Emotional states and physical health. American Psychologist 55(1), 110–121 (2000)
8. Lee, H., Choi, Y.S., Lee, S., Park, I.P.: Towards unobtrusive emotion recognition for affective social communication. In: IEEE Consumer Communications and Networking Conference, pp. 260–264 (2012)
9. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. IEEE Transactions on Patterns Analysis and Machine Intelligence 23, 1175–1191 (2001)
10. Ekman, P., Friesen, W.: The Facial Action Coding System. Consulting Psychologists Press, San Francisco (1978)
11. Devillers, L., Vasilescu, I., Lamel, L.: Emotion detection in a task-oriented dialog corpus. In: IEEE International Confererence on Multimedia & Expo (2003)
12. Gross, J.J., Barrett, L.F.: Emotion generation and emotion regulation: one or two depends on your point of view. Emotion Review 3(1), 8–16 (2011)
13. McRae, K., Misra, S., Prasad, A.K., Pereira, S.C., Gross, J.J.: Bottom-up and top-down emotion generation: implications for emotion regulation. Social Cognitive & Affective Neuroscience 7(3), 253–262 (2012)
14. Bachorik, J.P., Bangert, M., Loui, P., Larke, K., Berger, J., Roew, R., Schlaug, G.: Emotion in motion: investigating the time-course of emotional judgments of musical stimuli. Music Perception 26(4), 355–364 (2009)
15. Elliot, A.J., Maier, M.A.: Color and psychological functioning. Current Directions in Psychological Science 16(5), 250–254 (2007)
16. Pail, G., Huf, W., Pjrek, E., Winkler, D., Willeit, M., Praschak-Rieder, N., Kasper, S.: Bright-light therapy in the treatment of mood disorders. Neuropsychobiology 64, 152–162 (2011)
17. Wagner, J., Andre, E., Jung, F.: Smart sensor integration: a framework for multimodal emotion recognition in real-time. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–8 (2009)
18. Gehrig, T., Ekenel, H.K.: A common framework for real-time emotion recognition and facial action unit detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–6 (2011)
19. de Santos, A., Sánchez, C., Guerra, J., Bailador, G.: A stress-detection system based on physiological signals and fuzzy logic. IEEE Transactions on Industrial Electronics 58(10), 4857–4865 (2011)
20. Werner, K., Gross, J.J.: Emotion regulation and psychopathology: a conceptual framework. In: Emotion Regulation and Psychopathology, pp. 13–37. Guilford Press (2010)