Recognition Using Multi-Modal Data Fusion

Bahar Irfan, Tony Belpaeme
Centre for Robotics and Neural Systems
Plymouth University
Plymouth (UK)
bahar.irfan@plymouth.ac.uk, tony.belpaeme@plymouth.ac.uk

***Abstract*—In this paper, we introduce a multi-modal recognition method using Bayesian network. We employ face recognition as the primary identifier of a person and gender, age, height, time and location as the secondary identifiers, to increase the probability of correct recognition. The proposed network is to be used in two studies: the former for learning the optimal weights of the system through a HRI scenario, and the latter for use in cardiac rehabilitation for a personalised therapy assistant robot.**

I. Introduction

In social robotics, it is often essential to pursue an interaction with a robot for a long amount of time or over short, but consecutive interactions. One possible way to achieve a reliable long-term interaction between a robot and the user is through personalisation [1]. Wood et al [2] suggest that the adaptation of the agent requires semantic information such as remembering a name, and episodic information such as the commonly shared previous interactions with a user. As part of the APRIL project, we intend to create a memory system similar to a human memory, in that the user will be recognised and the salient information in the previous interactions will be remembered through episodic memory to achieve a better long-term interaction.

In terms of person recognition, face recognition is the most common method. However, most face recognition datasets use clear mug-shot images for learning and testing. During a real-time interaction, this would require the user to stay still for a few seconds and look directly into the camera, which is often not ideal. Furthermore, occlusions on the face such as glasses/sun glasses, or a different hair-style could effect the recognition results.

Person recognition made by humans, on the other hand, is not dependent on a single-modality. For example, let's take the scenario that someone is waving from afar at a person. This case would require a fast and correct reaction to the gesture. In that case, the person would initially think of the location and time of the day which could help eliminate possibilities. As the person approaches, the body size (i.e. height and the shape), clothing and hair style, gender and age could help distinguishing before seeing the face. Therefore, it would be possible to react before getting in a close distance to the person.

However, most of the times the primary biometric traits, such as the face, and the voice, are more reliable than the secondary biometric traits, as they only help eliminate other possibilities. Without seeing the face or hearing the voice, the recognition might be false as more than one person might match the combination of the secondary traits. Therefore, they should be used as supplementary to the primary traits, effect the decision less, and should not all have the same weight in the decision. Jain et al [3] demonstrate that using a Bayesian network for biometric recognition using fingerprints as primary biometric traits and gender, ethnicity, and height as 'soft' biometric traits improves the performance of recognition.

In this paper, we propose a similar system to [3], in that a Bayesian network is used to combine location, time, height, gender, age, with face recognition information. We intend to use this network for socially assistive robotics in cardiac rehabilitation, in order to recognise the patient and personalise the therapy [4]. The novelty of our work is in the application of Bayesian network in person recognition to HRI scenarios, and within the weighting system of the network.

II. Methodology

A. Bayesian Network

We use a Bayesian network in order to fuse the recognition data from multiple sources, which are face, gender, age, height, time and location. We use pyAgrum [5] library for implementing the Bayesian network structure.

As the Bayesian network structure requires, we use probabilities for each state within the variables. Each of these probabilities are normalised. We consider age, height and time as discrete random variables. In order to calculate the probabilities of all the possible states for these variables, we use the following approach. Given the system recognition confidence at a certain state, we consider that state as the mean of a discretised and normalised normal distribution. By knowing the mean of this probability mass function and the probability at that mean, we can estimate all the probabilities of the other states. For example, if a person is recognised as being 25 years old with probability 0.4, then we can find the probability of age 22 as 0.003 from the discretised and normalised normal distribution with mean 25 and standard deviation estimated as 0.953.

We use an "unknown" state in detecting the identity of the person, which corresponds to the person not being recognised either due to the person being unknown to the system or due to the maximum estimated probability being below the threshold.

We use weights for the secondary biometric traits (gender, age, height, time and location) in our system as Jain et al

[3] suggests. However, in our system the Bayesian network outcome is found by the chain rule, as opposed to the sum of logarithms method used in the paper. Therefore, we smooth the recognition results of each modality by using the weights as an exponent to the results.

SoftBank robotics recognition modules are used for input modalities [6]. We will be conducting two studies for validating our system in real-life HRI scenarios. The initial study uses a Pepper robot, in which the user interacts with the robot for initial registering and later confirming the recognition of the robot. The second study will use a Nao robot and external tablet for user input in cardiac rehabilitation therapy [4]. In both studies location is not used in the network, as the location within the experiments do not change.

B. Study 1: Registering and Recalling

This study aims to collect data for validating the system, and optimising the weights of the modalities in the network.

The robot interacts with the person detected for asking for confirmation of the estimated name or requests the name if the user cannot be recognised. In the latter case, if the entered name is not within the system, the user registers with the robot, which asks the gender, age, and height of the person, and takes a picture to learn the face.

C. Study 2: Cardiac Rehabilitation

In this study, we aim to analyse the effects of recalling the patient and their previous sessions in motivation during and after the therapy and in continuity of the therapy.

The experiments will be conducted in Instituto de Cardiologa at Fundacin Cardioinfantil clinic, in Bogota. The recommended cardiac therapy in the clinic is 18 weeks (36 sessions), and the study is intended to be continued for the full duration. Our study focuses on tracking heart rate, Borg scale (perceived exertion level) [7], cadence, step length, speed and gaze of the patient during the treadmill session in the therapy.

Three conditions will be used for this study: tablet condition, robot without personalisation and robot with personalisation. A total of 40 patients with similar diagnosis will participate in the study.

In the tablet condition, the sensor data is gathered, and a tablet requests for the Borg scale from the user with a beeping sound. No feedback is given to the user. In the robot without personalisation condition, the robot gives motivational feedback, validates the exertion level with the sensor data and alert the medical staff for changes when the heart rate or Borg scale exceeds the thresholds given by the medical staff, and also for medical condition input alert from the patients. In these two conditions, the user logs into the system using an identification number.

In addition to the functionalities described above, the robot with personalisation recognises the patient through our recognition network. Afterwards, the robot comments on the absence of the patient, if any, and refers to any problems in the previous session. Throughout the session, the robot calls the person by their name, and compares the current session with the previous one.

III. Discussion

The preliminary data from study 1 suggest that the network can be useful in predicting the person when one of the modalities fail, e.g. in case of a blurry image or when the person is not facing the camera. However, this could also lead to an unknown user being recognised as another person. Therefore, the weights of the parameters should be adjusted in order to decrease the false positive and the false negatives.

IV. Conclusion

Our studies are still ongoing, therefore, we have not achieved the final results yet. We will be comparing the results of the network to that of the face recognition of the built-in software in the robots, and to that of a state-of-art library such as Openface [8].

Currently feedback in the second study is based upon the criteria, such as the period for requesting Borg scale and the critical thresholds, that are estimated by the medical staff. Through the data gathered in the second study, we intend to analyse the patterns in therapy in order to extract the differences and similarities in the patients to find the ideal thresholds for cardiac rehabilitation therapy.

V. Acknowledgments

This work has been supported by the EU H2020 Marie Sklodowska-Curie Actions APRIL project (grant 674868).

References

[1] K. Dautenhahn, “Robots We Like to Live With ?! - A Developmental Perspective on a Personalized , Life-Long Robot Companion,” in *2004 IEEE International Workshop on Robot and Human Interactive Communication*, 2004, pp. 17–22.

[2] R. Wood, P. Baxter, and T. Belpaeme, “A review of long-term memory in natural and synthetic systems,” *Adaptive Behavior*, vol. 20, no. 2, pp. 81–103, 2011.

[3] A. K. Jain, S. C. Dass, and K. Nandakumar, “Soft biometric traits for personal recognition systems,” in *International Conference on Biometric Authentication*, ser. LNCS, no. 3072, July 2004, pp. 731–738.

[4] J. S. Lara Ramirez, J. A. Casas Bocanegra, A. F. Aguirre Fajardo, M. Munera, M. Rincon-Roncancio, B. Irfan, E. Senft, T. Belpaeme, and C. A. Cifuentes Garcia, “Soft biometric traits for personal recognition systems,” in *IEEE-RAS-EMBS International Conference on Rehabilitation Robotics*, 2017.

[5] C. Gonzales, L. Torti, and P.-H. Wuillemin, “aGrUM: a Graphical Universal Model framework,” in *International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems*, ser. Proceedings of the 30th International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems, Arras, France, Jun. 2017. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01509651

[6] SoftBank Robotics. Naoqi documentation. [Online]. Available: http://doc.aldebaran.com/2-4/index.html

[7] G. Borg, “Perceived exertion as an indicator of somatic stress,” *Scand J Rehabil Med.*, vol. 2, no. 2, pp. 92–98, 1970.

[8] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, “Openface: A general-purpose face recognition library with mobile applications,” CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.