# Robots Should Respect Your Feelings as Well

## Adapting Distance Between a Robot and an User Based on Expressed Emotions


Markus Bajones
Automation and Control Institute
TU Wien
Vienna, Austria
Email: markus.bajones@tuwien.ac.at

Michael Zillich
Automation and Control Institute
TU Wien
Vienna, Austria
Email: michael.zillich@tuwien.ac.at

Markus Vincze
Automation and Control Institute
TU Wien
Vienna, Austria
Email: markus.vincze@tuwien.ac.at



***Abstract*—As robots move out of closed and controlled facilities and into domestic environments, these robots need to observe and understand the surrounding world it is part of. Creating behavior a user expects is an important step to create successful human-robot interaction in the long run. Especially the way a mobile robot moves towards a person in a one-on-one situation sets expectations for later interactions. Commonly this is only based on the distance between the person and the robot. The emotions humans show when a robot moves in their close proximity or the amount of attention a person is giving towards the robot however, have rarely been considered in real time operations. In this work we present an emotion and attention recognition pipeline of a simple robotic behavior of adapting the distance to a human based on the capabilities of the installed sensors (field of view, range, etc.) as well as the emotions that the robot is observing from the humans facial expressions.**


## I. Introduction

Traditionally robots have been placed in enclosed working areas to ensure the safety of human workers. This approach, however toned down nowadays, is still often used in an industrial setting. In other environments and situations (e.g. costumer support robot in a shop, care robot in a care facility) this approach is neither possible nor purposeful. Substantial progress has been made to enable robots to keep at an appropriate distance to the user based on personal zones (proxemics) as defined by Hall [1], or on the robot's sensor requirement to be within a certain range [2]. Dondrup et al. [3] incorporated Qualitative Trajectory Calculus into the concept of proxemics during interactions of a mobile robot and a person. We believe however that social, mobile robots need to adapt their behavior not only based on the distance, but also according to expressed emotions of the people they are interacting with. To enable such adaptation we propose a system that classifies an emotion from a RGB image into one out of six basic emotions (anger, disgust, fear, happiness, sadness, and surprise). Further it calculates the duration of visual attention (i.e. looking at the robot) given towards the robot. This is used as an indicator of the awareness of the robot's actions by the user. Recent work on facial emotion recognition moved away from the focus of still images towards video sequences and shows promising results in both fields [4], [5]. The step to put such systems on a mobile robot however not only brings challenges in terms of a fast moving camera (e.g. on a pan/tilt unit, or arm) but also the ability to reposition the robot to the user for improved recognition rates.

## II. Proposed Architecture

To extend a mobile robot's capabilities in domestic environment or public space we designed an emotion recognition system working on RGB images. While sensor input from e.g. heart rate monitor or EEG could improve the classification of emotions, we do not see them as a practical approach outside of a lab situation. Therefore, we limit the robot's input data to reflect what a human would be able to observe as well.

In the following we only present the steps starting after the image acquisition. The detection of the human based on LIDAR distance measurements and viewpoint calculations to point the camera towards the face of the user is not discussed. An overview of this pipeline is also presented in figure 1.

- The first step is the commonly used face- and facial landmark detector from the DLib toolkit[1]. The output from this detection is fed into 3 parallel blocks for emotion recognition and gaze direction calculations.
- For the next block we retrained the final layers of Google's Inception network [6] to classify the basic emotions based on facial expressions. The retraining was performed on the FER-2013 facial expressions dataset, created by Pierre Luc Carrier and Aaron Courville as outlined in [7]. As the training of the original inception network was not performed with facial landmarks in mind we do not use them in this block and only operate on the region-of-interest (ROI) in which a face was detected.
- For the next emotion recognition block we trained a more traditional machine learning approach, namely a SVM classifier. We trained this classifier on the same dataset as our Inception based method, however instead of a ROI we directly used the facial landmarks extracted from the first step. The landmarks were extracted from the same dataset as for our first method.
- Further we added a block to calculate a head pose- and gaze direction estimation based on the work from [8] which we use to infer the amount of attention a person is giving towards the robot.

Parallel to the emotion recognition we collect the duration of attention a user is providing towards the robot and the

[1]https://dlib.net

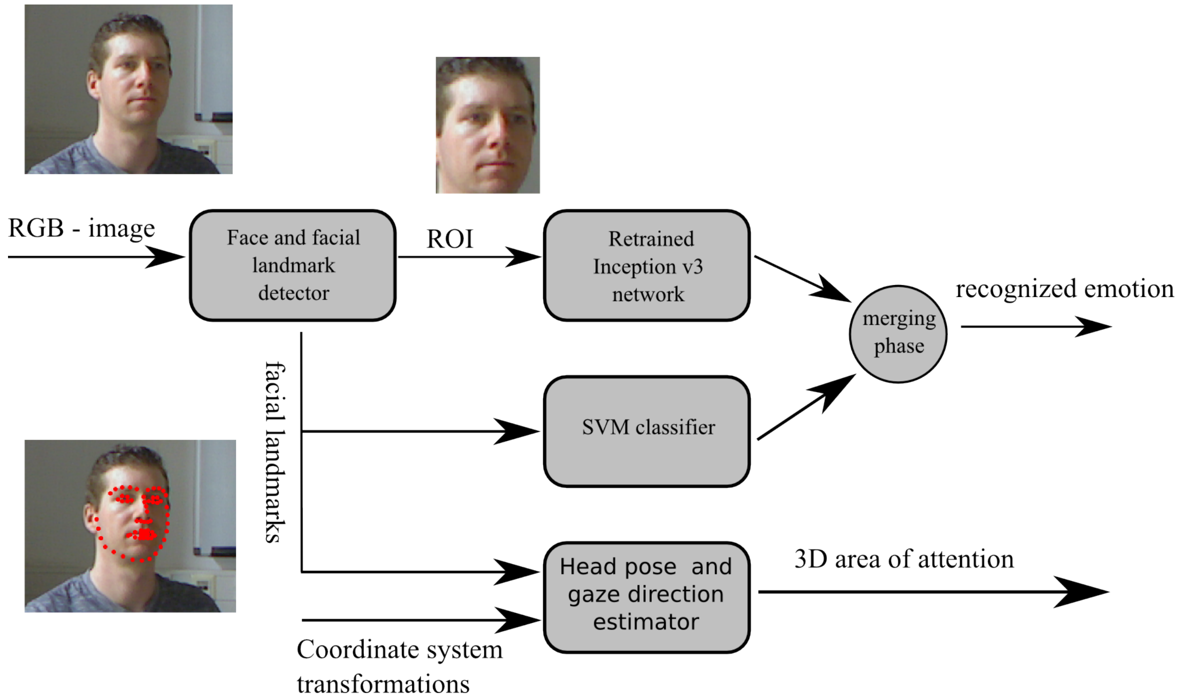


Fig. 1. Emotion recognition and attention estimation pipeline

last recognized emotion to enable the robot's decision making framework to incorporate into. In our case we plan to explore this to adapt its distance towards the user in different interaction scenarios.

## III. Discussion

First tests with an Asus Xtion Pro sensor mounted on the pan/tilt unit on top of a Festo Robotino2 platform suggest that the optimal distance for the face detection and landmark extraction is between $0.9\,\mathrm{m}$ and $1.9\,\mathrm{m}$ from the camera. This range will be used to define the area in which our robot should stay during interaction.

Our next step is to collect data on expectation towards a mobile robot based on emotional state. From this we expect a first simple model about the movements the robot should perform after recognizing a certain emotional state of the user. After that we are planning a user study in which we will evaluate this model on two different mobile robots to evaluate not only the effect of the adaptation of distance, but also if such an effect can be observed across different robotic embodiments as well. Our overall expectation is that such an adapting behavior will result in higher acceptance of the robot, as it adds a sense of situation awareness into the overall robotic system.

## IV. Acknowledgments

The research leading to the presented results has received funding from the European Communitys Seventh Framework Programme under grant agreement No. 610532 - SQUIRREL and Austrian Research Promotion Agency (FFG) under grant agreement 848653 - AssistMe.

## References

[1] Hall and E. Twitchell, *The hidden dimension, 1st ed.* Doubleday & Co, 1966. [Online]. Available: http://psycnet.apa.org/psycinfo/2003-00029-000

[2] R. Mead and M. J. Mataric, "Robots Have Needs Too: People Adapt Their Proxemic Preferences to Improve Autonomous Robot Recognition of Human Social Signals," *Human-Robot Interaction*, vol. 5, no. 2, pp. 48–68, 9 2016. [Online]. Available: https://www.cs.kent.ac.uk/events/2015/AISB2015/proceedings/hri/19-Mead-robotshaveneeds.pdf

[3] C. Dondrup, N. Bellotto, M. Hanheide, K. Eder, and U. Leonards, "A Computational Model of Human-Robot Spatial Interactions Based on a Qualitative Trajectory Calculus," *Robotics*, vol. 4, no. 1, pp. 63–102, 3 2015. [Online]. Available: http://www.mdpi.com/2218-6581/4/1/63/

[4] M. H. Siddiqi, F. Farooq, and S. Lee, "A robust feature extraction method for human facial expressions recognition systems," *Proceedings of the 27th Conference on Image and Vision Computing New Zealand - IVCNZ '12*, pp. 464–468, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2425836.2425924

[5] Z. Yu and C. Zhang, "Image based Static Facial Expression Recognition with Multiple Deep Network Learning," *ACM on International Conference on Multimodal Interaction - ICMI*, pp. 435–442, 2015. [Online]. Available: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/icmi2015_ChaZhang.pdf

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper With Convolutions," pp. 1–9, 2015. [Online]. Available: http://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html

[7] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in Representation Learning: A report on three machine learning contests," *ArXiv e-prints*, 2013.

[8] S. Lemaignan, F. Garcia, A. Jacq, and P. Dillenbourg, "From Real-time Attention Assessment to With-me-ness in Human-Robot Interaction," in *Proceedings of the 2016 ACM/IEEE Human-Robot Interaction Conference*, 2016. [Online]. Available: http://github.com/severin-lemaignan/gazr