2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin).

Visual-based sentiment logging in magic smart mirrors

Simone Bianco, Luigi Celona, Paolo Napoletano
Department of Informatics, Systems and Communication
University of Milano-Bicocca
viale Sarca, 336, 20126, Milano, Italy
Email: {bianco, luigi.celona, napoletano}@disco.unimib.it

***Abstract*—This paper describes the hardware and software architectures of a smart magic mirror able to acquire and track the user's face, recognize his identity, analyze and log his facial expressions and emotional states. The magic mirror is basically a see-through mirror made smart by a led display placed behind the mirror that enables to display the User Interface (UI). The mirror is connected to a small single-board computer attached to a set of input sensors (a traditional RGB camera to enable vision-based interaction, a microphone to enable voice interaction, temperature and humidity sensors, and proximity sensors) and to an embedded machine intelligence platform that performs all the neural computations.**

***Index Terms*—Magic smart mirror, Face sentiment analysis, Convolutional Neural Networks, Internet of Things, Embedded Systems, Sentiment logging.**

I. Introduction

Recently, different companies and research laboratories have started developing smart mirrors, which are mainly used to display information in smart home scenarios or inside shops. Most of them are developed just to show information such as time,weather, products catalogs, etc. More sophisticated ones have one or more cameras to acquire the subjects and display blended reality views [1], [2]. Open source modular smart mirror platforms also exists (e.g. MagicMirror[2]) and have a huge number of data visualization modules available. However, they lack the modules for acquiring and analyzing the user.

In this work, we propose a magic smart mirror that acquires and tracks users' faces, analyzes and logs theirs facial expressions and emotional states.

The proposed magic smart mirror has three different states. The first one is the off-state, in which it behaves like a standard mirror. The second one is the stand-by status: this state is reached when the mirror detects the presence of human(s) in front of it using proximity sensors and switches it off when the user goes away; also in this state the mirror behaves like a standard mirror. From the stand-by the on-state can be activated by vocal command; in this state all the smart functionalities of the mirror are activated and the mirror displays the User Interface (UI) and the chosen widgets/modules. We have introduced in our prototype some modules related to face recognition and analysis that can be optionally displayed. This paper focuses on the design and implementation of these novel face-related features.

II. Hardware architecture

Fig. 1 depicts the general architecture of the our magic smart mirror. The entire system is based on the Raspberry PI 3[1] that is a small single-board computer developed by the Raspberry Pi Foundation. The magic mirror effect is obtained by using an acrylic see-through mirror placed on the top of a consumer monitor then connected through the HDMI cable to the Raspberry board. A digital temperature and humidity sensor as well as a proximity sensor are connected through the GPIO interface to the Raspberry. The left and right speakers are connected through the 4-pole TRRS output. The audio output is alternatively obtained directly from the monitor through the HDMI video signal. The camera used for face recognition is the 8Mp Raspberry Pi Official Camera Module V2[2] connected through the CSI cable extender to the CSI input of the board. The neural processing is done through the Intel Movidius Neural Compute Stick[3] (NCS) that is suitable designed to handle Caffe models [3]. The stick is connected through the USB port to the board. Finally, a microphone is connected through the USB port to the board.

III. Software architecture

The software architecture of our system is implemented through the popular web socket communication protocol [4]. The architecture is modular and each single module can be easily integrated in the MagicMirror[2] platform. This platform allows to easily implement a general purpose smart magic mirror on a Raspberry PI computer board with some basic functionalities, such as clock, calendar, current weather, etc, and many other third-party functionalities. We designed our functionalities with the aim to be easily integrated in the next future within the MagicMirror[2] platform as third-party module.

Fig. 2 shows the software architecture of our system. The server module leads all the operations: interaction with the user through the sensors and the mirror user interface (UI), information visualization through the mirror UI, neural computation with the face recognition and emotion recognition, data management through the interaction with the SQLite3

[1]https://www.raspberrypi.org/
[2]https://www.raspberrypi.org/products/camera-module-v2/
[3]https://developer.movidius.com/

978-1-5386-6095-9/18/$31.00 ©2018 IEEE

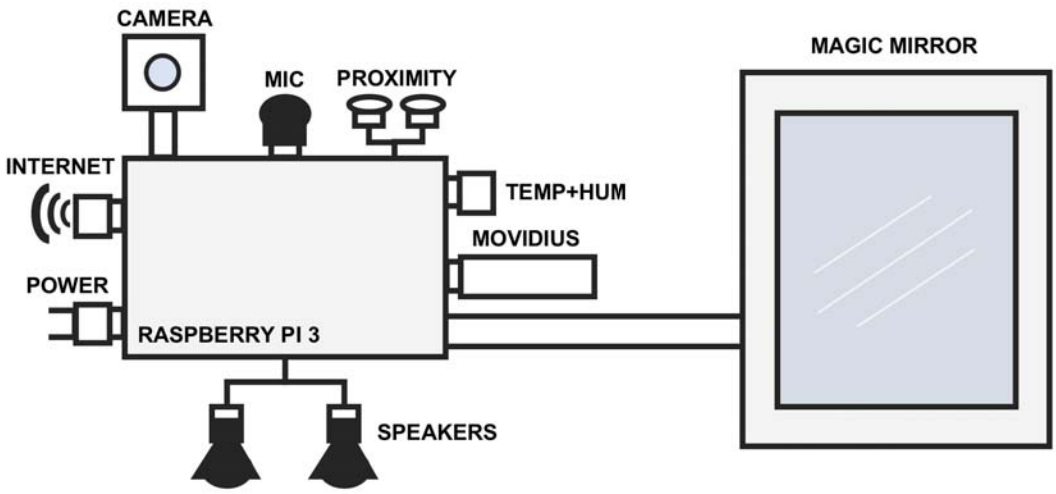

Fig. 1. Hardware architecture of our magic smart mirror.

database. Summing up, the server module operates four main functionalities following a multi-thread paradigm: *world sensing*, *information processing*, *information management* and *information visualization*.

A. World sensing

The magic smart mirror has three different states:

1) **off-state:** the mirror behaves like a standard mirror.
2) **stand-by-state**: the mirror is ready to be activated. The mirror, through its proximity sensor and microphone, senses the surrounding environment and detects the presence of human(s) in front of the mirror.
3) **on-state**: the mirror reacts to the surrounding environment by recognizing human identities and monitoring emotions.

The world sensing thread is designed to acquire and handle all the signals coming from the input sensors:

- video signal from the camera: this signal is used by the neural module;
- audio signal from the microphone: this signal is used by the neural module;
- temperature and humidity signals from the corresponding sensor: this signals are used by the information visualization module;
- proximity signal from the corresponding sensor: the analysis of this signal allows to change the state of the mirror from stand-by-state to on-state.

B. Information processing through the neural module

1) Speech recognition and synthesis: Automatic Speech Recognition (ASR) is performed using the Google Cloud speech-to-text API. The service can stream text results, immediately returning text as it is recognized from streaming audio or as the user is speaking. For what concerns speech synthesis, this is performed using the Google Cloud text-to-speech API, which enables to synthetize natural-sounding speech with different voices, available in multiple languages and variants.

2) Face detection and recognition: The magic smart mirror leverages face recognition technology to match the user's face to their profile. The user creates a profile the first time his face is detected by the mirror and is not recognized as an existing user. The face detector used is a Convolutional Neural Network (CNN) based object detector trained using the Max-Margin Object Detection loss [5].

Once the face is detected, 68 face landmarks are detected using an ensemble of randomized regression trees that regress the location of facial landmarks from a sparse subset of intensity values extracted from the input image [6]. The position of these facial landmarks is used to fit a geometric transform that aligns the detected face to a reference canonical frontal pose.

From the aligned face of the new user is then computed a unique face identifier (i.e. a face-id) which is then stored in the mirror's database. The face identifier is computed by extracting the activations of a CNN trained to recognize a large number of different identities enforcing inter-class dispersion and intra-class compactness [7], [8], while at the same time recognizing identities across large variations in face appearance [9].

3) Face attributes: The magic smart mirror uses the multi-task learning approach based on convolutional neural network (MTL-CNN) to jointly estimate multiple facial attributes from a single face image proposed in [10]. While other methods centered on single attributes/expression exist [11], the MTL-CNN model takes into account the attribute inter-correlations to obtain informative and robust feature representation. The model simultaneously predicts one among 8 disjoint age groups (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+ years) as available in the Adience benchmark [12], gender, and 40 binary attributes as available in the CelebA dataset [13] such as: sideburns, arched eyebrows, brown hair, smiling, attractive, straight hair, bags under eyes, narrow eyes, wavy hair, bald, no beard, wearing earrings, eyeglasses, etc.

4) Face expressions and emotional state recognition: The magic smart mirror is able to recognize human affects through the analysis of facial behavior. Specifically, given a single face image, the proposed MTL-CNN simultaneously predicts:

- One discrete emotional state among the eight "basic", namely neutral, happiness, sadness, surprise, anger, fear, disgust, and contempt; and other expressions, different from the previous eight, such as sleepy, bored, tired, confused, disgusted, pleased, etc.;
- Two continuous emotional scores in the range $[-1, 1]$ for valence and arousal respectively.

The proposed model is trained on the Affect from the interNet (AffectNet) database [14], which is the largest collection of facial images for facial expressions, valence and arousal in the wild. All the different discrete and continuous emotional states can be plotted together in a plot similar to the one reported in Fig. 3: the discrete emotional states are reported outside the circumference, with each of one of them being characterized by a different angle; the continuous emotional scores are used to identify the intensity of the detected emotion, ranging from low intensity when valence and arousal have both a small magnitude to high intensity when at least one of them has a large magnitude.

5) Sentiment logging and short-term/long-term analysis: The magic mirror performs a real-time continuous estimation of face expression and emotional state, and logs them together with the corresponding user's identity. The current estimate is displayed to the user as reported in Fig. 4(a), thus permitting a short-term analysis. All the user's logs in a selected time range (e.g. one day, one week, one month, etc.) are used to perform a long-term analysis by computing the emotion distribution. The report is displayed as reported in Fig. 4(b)

C. Information management

The mirror logs sensors data and information processed by the neural module. In particular, the information logged are:

- daily emotions;
- daily temperature and humidity values;
- daily mirror usage;
- face recognition descriptors.

This information is logged in a SQLite3 database.

D. Information visualization

The mirror displays daily temperature and humidity, weekly statistics of emotional state and current emotional state through a web interface. The web interface is enriched with the d3.js library that is a JavaScript visualization library [15].

IV. CONCLUSION

In this paper we presented the hardware and software architectures of a smart magic mirror that using a traditional RGB camera is able to acquire and track the user's face, recognize his identity, analyze and log his facial expressions and emotional states permitting a short-term and long-term analysis.

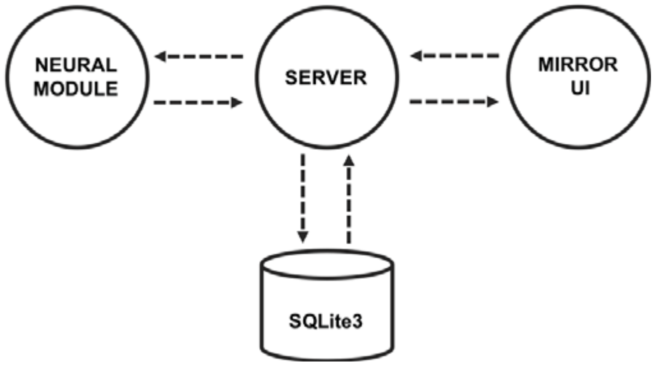

Fig. 2. Software architecture of our magic smart mirror.

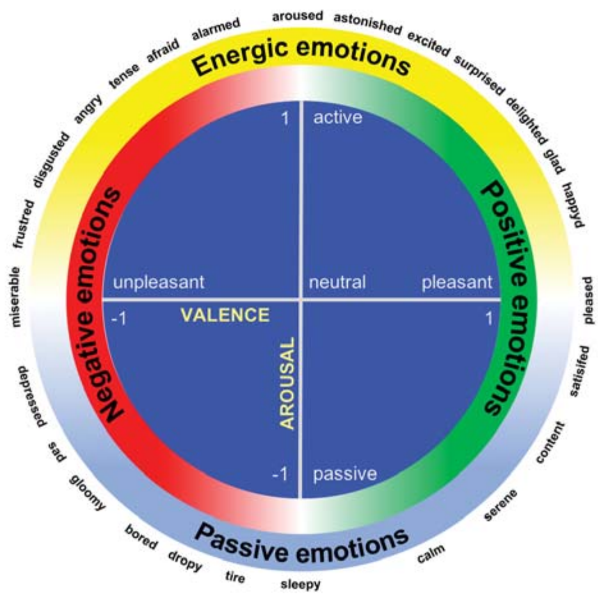

Fig. 3. Visualization interface for the face expressions and emotional state recognition: the discrete emotional states are reported along the circumference, while the predicted continuous values for valence and arousal are plotted inside the circle. Additional information about negative vs positive emotions and energic vs passive emotions are also reported to increase the interpretability of the results.

The mirror is connected to a Raspberry single-board computer, that is attached to a set of input sensors (a camera, a microphone, temperature and humidity sensors, and proximity sensors) and to the Intel Movidius Nural Compute Stick, which is an embedded machine intelligence platform that performs all the neural computations.

As future works we plan to increase the number of cameras used, in order to use the magic smart mirror for Augmented Reality (AR) applications, and to connect it with a larger number of heterogeneous sensors (such as physiological sensors) and thus being able also to monitor human stress [16]. We also plan the integration of the developed modules in the MagicMirror[2] open source modular smart mirror platform.

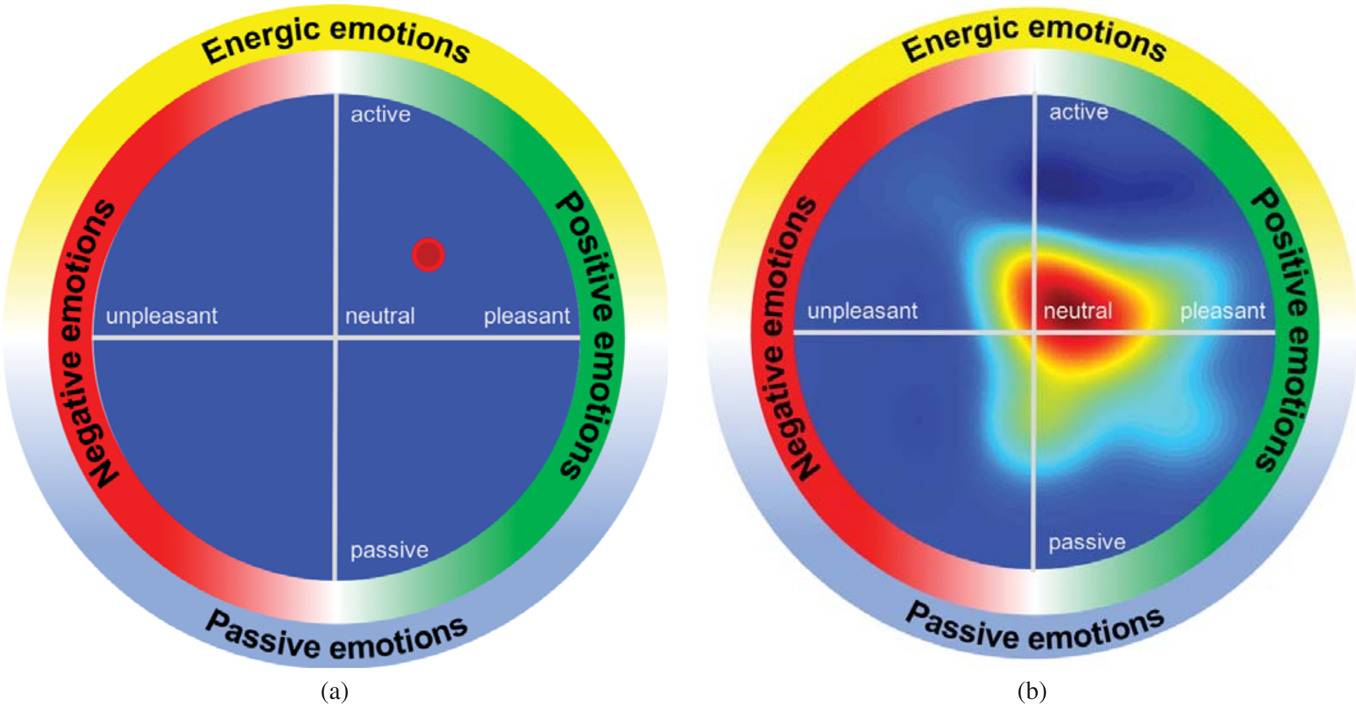

Fig. 4. Visualization interfaces for the face expressions and emotional state recognition to perform short-term (i.e. real-time) analysis (a) and long-term analysis (b).

ACKNOWLEDGMENTS

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. This research leading to these results has received funding from The Home of Internet of Things (Home IoT), CUP: E47H16001380009 - Call "Linea R&S per Aggregazioni" cofunded by POR FESR 2014-2020. We thank Dr. Marco Fotino for his help in developing the mirror prototype.

REFERENCES

[1] T. Blum, V. Kleeberger, C. Bichlmeier, and N. Navab, "mirracle: An augmented reality magic mirror system for anatomy education," in *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*. IEEE, 2012, pp. 115–116.

[2] I. Amazon Technologies, "Blended reality systems and methods," 2018, uS Patent 9,858,719.

[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[4] HTML Standards, "Web sockets," https://html.spec.whatwg.org/multipage/web-sockets.html, 2018, [Online; accessed 1-June-2018].

[5] D. E. King, "Max-margin object detection," *arXiv preprint arXiv:1502.00046*, 2015.

[6] V. Kazemi and S. Josephine, "One millisecond face alignment with an ensemble of regression trees," in *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*. IEEE Computer Society, 2014, pp. 1867–1874.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[8] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 499–515.

[9] S. Bianco, "Large age-gap face verification by feature injection in deep networks," *Pattern Recognition Letters*, vol. 90, pp. 36–42, 2017.

[10] L. Celona, S. Bianco, and R. Schettini, "Fine-grained face annotation using deep multi-task cnn," *Sensors (submitted)*, 2018.

[11] S. Bianco, L. Celona, and R. Schettini, "Robust smile detection using convolutional neural networks," *Journal of Electronic Imaging*, vol. 25, no. 6, p. 063002, 2016.

[12] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[13] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3730–3738.

[14] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *Transactions on Affective Computing*, 2017.

[15] D3JS, "JavaScript library," https://d3js.org/, 2018, [Online; accessed 1-June-2018].

[16] P. Napoletano and S. Rossi, "Combining heart and breathing rate for car driver stress recognition," in *International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*. IEEE, 2018, pp. –.