

## Central Lancashire Online Knowledge (CLOK)

Title	Mirror mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization
Type	Article
URL	<a href="https://clock.uclan.ac.uk/id/eprint/17036/">https://clock.uclan.ac.uk/id/eprint/17036/</a>
DOI	<a href="https://doi.org/10.1109/TMM.2017.2666545">https://doi.org/10.1109/TMM.2017.2666545</a>
Date	2017
Citation	Henriquez Castellano, Pedro, Matuszewski, Bogdan J., Andreu-Cabedo, Yasmina, Bastiani, Luca, Colantonio, Sara, Coppini, Giuseppe, D'Acunto, Mario, Favilla, Riccardo, Germanese, Danila et al (2017) Mirror mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization. IEEE Transactions on Multimedia, 19 (7). pp. 1467-1481. ISSN 1520-9210
Creators	Henriquez Castellano, Pedro, Matuszewski, Bogdan J., Andreu-Cabedo, Yasmina, Bastiani, Luca, Colantonio, Sara, Coppini, Giuseppe, D'Acunto, Mario, Favilla, Riccardo, Germanese, Danila, Giorgi, Daniela, Marraccini, Paolo, Martinelli, Massimo, Morales, Maria-Aurora, Pascali, Maria Antonietta, Righi, Marco, Salvetti, Ovidio, Larsson, Marcus, Stromberg, Tomas, Randeberg, Lise, Bjorgan, Asgeir, Giannakakis, Giorgos, Pediaditis, Matthew, Chiarugi, Franco, Christinaki, Eirini, Marias, Kostas and Tsiknakis, Manolis

It is advisable to refer to the publisher's version if you intend to cite from the work.  
<https://doi.org/10.1109/TMM.2017.2666545>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLOK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

# Mirror mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization

Pedro Henriquez<sup>1</sup>, Bogdan J. Matuszewski<sup>1</sup>, Yasmina Andreu<sup>1</sup>, Luca Bastiani<sup>2</sup>, Sara Colantonio<sup>2</sup>, Giuseppe Coppini<sup>2</sup>, Mario D'Acunto<sup>2</sup>, Riccardo Favilla<sup>2</sup>, Danila Germanese<sup>2</sup>, Daniela Giorgi<sup>2</sup>, Paolo Marraccini<sup>2</sup>, Massimo Martinelli<sup>2</sup>, Maria-Aurora Morales<sup>2</sup>, Maria Antonietta Pascali<sup>2</sup>, Marco Righi<sup>2</sup>, Ovidio Salvetti<sup>2</sup>, Marcus Larsson<sup>3</sup>, Tomas Strömberg<sup>3</sup>, Lise Randeberg<sup>4</sup>, Asgeir Bjorgan<sup>4</sup>, Giorgos Giannakakis<sup>5</sup>, Matthew Pediaditis<sup>5</sup>, Franco Chiarugi<sup>5</sup>, Eirini Christinaki<sup>5</sup>, Kostas Marias<sup>5</sup>, Manolis Tsiknakis<sup>5,6</sup>

**Abstract**—A person's well-being status is reflected by their face through a combination of facial expressions and physical signs. The SEMEOTICONS project translates the semeiotic code of the human face into measurements and computational descriptors that are automatically extracted from images, videos and 3D scans of the face. SEMEOTICONS developed a multisensory platform in the form of a smart mirror to identify signs related to cardio-metabolic risk. The aim was to enable users to self-monitor their well-being status over time and guide them to improve their lifestyle. Significant scientific and technological challenges have been addressed to build the multisensory mirror, from touchless data acquisition, to real-time processing and integration of multimodal data.

**Index Terms**—Cardio-metabolic risk, unobtrusive well-being monitoring, multimodal data integration, 3D face detection and tracking, 3D morphometric analysis, psychosomatic status recognition, multispectral imaging, breath analysis.

## I. INTRODUCTION

THE principal communication channel among humans is the face; it is a mirror of physical conditions, mood and emotions. As such, the face is the basis of medical semeiotics, revealing the well-being status of an individual through facial expressions and a combination of physical signs (e.g., subcutaneous fat, skin color). This paper describes how the EU FP7 project SEMEOTICONS (<http://www.semeoticons.eu/>) moves medical semeiotics to the digital realm, translating the semeiotic code of the face into measurements and computational descriptors obtained from images, videos and 3D scans of the human face. The developed Wize Mirror, an intelligent multisensory device, can detect and monitor facial signs over time correlating them with cardio-metabolic risk and providing personalized guidance to users on how to improve their habits. Cardiac-related conditions are the leading cause of mortality worldwide, therefore, a device that can monitor cardio-metabolic risk is an important tool to maintain a healthy lifestyle.

Other smart mirrors have been proposed for different purposes, such as virtual clothes fitting [1], make-up rendering [2] and biofeedback [3]. The latter paper described a multimodal system using an interactive mirror and biomedical sensors (camera, hand-held ECG, blood pressure and skin temperature sensors and pressure pad). The main difference between [3] and the device developed in this work is that the Wize Mirror integrates a user-friendly interface (Fig. 1) with breath composition analysis and contactless imaging of facial features for advanced multimodal physiological analysis. The different sensors collect heterogeneous data, including (multispectral) images, videos, 3D scans and gas concentration signals, from the user in front of the mirror. Dedicated algorithms process the data, and extract colorimetric, morphometric, biometric and compositional descriptors of facial signs. According to the semeiotic model of the face for cardio-metabolic risk [4], the descriptors include:

- 3D morphological face descriptors related to asymmetry, swelling, overweight and obesity computed from a 3D reconstruction of the face (Section IV).
- Facial descriptors revealing emotional status including stress, anxiety and fatigue, captured via 2D expression analysis on video sequences (Section V).
- Physiological parameters such as respiratory rate, heart rate and heart rate variability, all estimated from videos by detecting subtle color changes and cyclic movement during the observation time (Section VI).
- Descriptors associated with cholesterol, metabolic end products found in diabetes and endothelial dysfunction, evaluated using a novel multispectral imaging system assessing skin tissue, including microcirculation (Section VII).
- Exhaled gas composition, measured using a novel gas-sensing device, which gives quantitative feedback about noxious habits such as smoking and alcohol intake (Section VIII).

All these descriptors are integrated to define a virtual individual model and an individual wellness index. This index enables a user to self-monitor and self-assess their well-being status over time. The Wize Mirror also offers personalized user guidance towards the achievement and maintenance of a

<sup>1</sup>University of Central Lancashire, Preston, UK

<sup>2</sup>National Research Council of Italy, Pisa, Italy

<sup>3</sup>Linköping University, Linköping, Sweden

<sup>4</sup>Norwegian University of Science and Technology, Trondheim, Norway

<sup>5</sup>Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Greece

<sup>6</sup>Department of Informatics Engineering, Technological Educational Institute of Crete

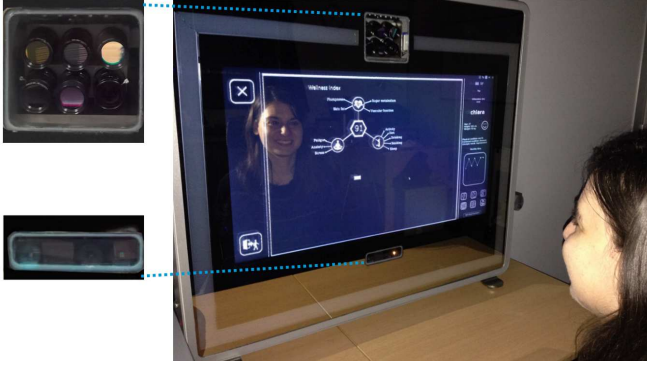


Fig. 1. The Wize Mirror prototype; on the left the multisensory rack, with from top to bottom: 5 multispectral imaging cameras (MSI), and a 90 fps color video camera (CV) and depth sensor (DS).

positive lifestyle.

Designing and building the Wize Mirror required solving significant scientific and technological challenges, from touchless data acquisition to real-time processing of multimodal data to obtain reliable measurements correlated with clinical risk factors. Section II introduces the sensing modalities used by the Wize Mirror. It also briefly describes how the implemented processing workflow integrates all these multimodal data with the calculated cardio-metabolic risk descriptors and the multimedia interface enabling ubiquitous and unobtrusive user interactions. The taxonomies, describing the data processing flow and the data sources, are introduced to guide the reader through the intricacy of the system detailed in the subsequent sections. Section III describes the processing multimedia backbone of the system, that is used by the different Wize Mirror sensors for user detection, recognition, positioning, labelling and 3D measurement. The details of each multimodal processing subsystem are presented in Sections IV-VIII. Finally, the definition of the virtual individual model integrating all calculated risk descriptors is presented in Section IX and the conclusions are described in Section X.

## II. MULTIMODAL DATA INTEGRATION

Different data fusion and multi-sensor integration tasks are performed on the Wize Mirror, including blending of 3D and 2D inputs, merging multispectral images and building the virtual individual well-being model.

This section briefly describes the sensing modalities used in the mirror and how they correspond to different semeiotic model descriptors. The sensors' interaction structure and overall dataflow are also presented to aid the description of the system integration strategy. The details of each sensing modality and the corresponding data analysis method are described in the subsequent sections. The multiple sensors and computed descriptors are given in Table I. Additionally, the table lists the sections of the paper where the details of the corresponding data analysis are given.

A multi-sensor integration is performed using 3D data (obtained from the depth sensor) to simplify the face detection and labelling process on the 2D data from the other cameras. This makes the system more robust and efficient. A common

TABLE I  
THE WIZE MIRROR SENSORS AND THE SEMEIOLOGICAL MODEL DESCRIPTORS  
CALCULATED BASED ON THE CORRESPONDING SENSING MODALITY.

Sensor	Descriptor	Section
Depth Sensor (DS)	3D Face Morphology (3DFM)	IV
Color Video (CV)	Facial Descriptors (FaDe)	V
Color Video (CV)	Physiological Parameters (PhPa)	VI
Multispectral Imaging Cameras (MSI)	Multispectral Measurements (MuMe)	VII
Wize Sniffer (WS)	Exhaled Gas Compositions (ExGC)	VIII

stage for the analysis is face detection, tracking and facial landmark estimation (see Fig. 3). Face detection and tracking is only performed once using the data from the depth sensor rather than on each data stream. This approach is more robust to varying illumination. In this method, the user is detected first in 3D space, and a model face mask is matched to the depth sensor data to estimate the position and orientation of the user's face. Subsequently, the positions of predefined facial landmarks are calculated (see Section III). Then, the 3D face coordinates are projected into the 2D frames of different cameras using the intrinsic and extrinsic parameters of the cameras. These parameters are obtained when the mirror is assembled using camera and stereo calibration techniques ([5] and [6]). Using the proposed workflow, the face detection and 2D facial landmark estimation stage is reduced to a single process using the depth data and six projections of the matched 3D face mask into the corresponding 2D image space (see Fig. 2). The projection is performed by multiplying each 3D vertex, given in homogeneous coordinates, of the matched face mask by a camera projection matrix  $M = A[R|T]$  built from 3x4 extrinsic parameter matrix (encoding mask rotation  $R$  and translation  $T$ ) and 3x3 matrix  $A$  describing the intrinsic camera parameters:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = A [R|T] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

where  $u/w$  and  $v/w$  represent the 2D coordinates in the image reference systems;  $x, y$  and  $z$  are the 3D point coordinates from the depth sensor reference system;  $R$  is a 3x3 rotation matrix encoding the rotation about three axes;  $T$  is a vector containing the translation from the origin and:

$$A = \begin{pmatrix} f & \alpha & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

where  $f$  is the focal length;  $\alpha$  is the skew parameter; and  $(u_0, v_0)$  are the coordinates of the principal point.

The computational descriptors of face signs (given in Section I) are integrated into the virtual individual model to build a representation of the users status, which is consistent with cardio-metabolic risk (see Fig. 3). Data fusion techniques are used to synthesize the wellness index, a non-diagnostic

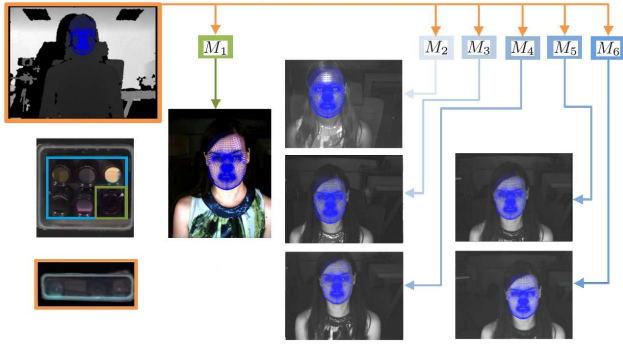


Fig. 2. Face detection and tracking performed on depth data and projected into multiple 2D images from different cameras. The depth sensor is highlighted in orange (DS), color video camera is highlighted in green (CV) and the five multispectral imaging cameras are shown in blue (MSI). Small colored boxes on top ( $M_1$  to  $M_6$ ) represent different camera projection matrices calculated using camera calibration. They are applied to the detected face point coordinates given in the depth sensor reference system to convert them into each camera's reference system.

estimation for self-assessment and self-monitoring of cardio-metabolic risk. Conceptually, the values obtained from the analysis can be seen as the components of a state vector moving in a multidimensional well-being state space. Subsequently, the model is mapped into three separate wellness subspaces related to physical wellness, emotional wellness and lifestyle habit wellness (see Section IX for details).

To facilitate validation of the Wize Mirror subsystems and as part of the mirror development a dedicated SEMEOTICONS' Reference Dataset (*SRD*) was built. The *SRD* includes all the modalities described in this paper as well as the results of reference clinical tests capturing 46 different physiological parameters. Additionally, 12 behavioral and psychometric parameters were obtained using clinically validated questionnaires. The reference dataset was collected in two acquisition campaigns conducted in May 2014 (*SRD'14*) and May 2015 (*SRD'15*) at the National Research Council of Italy in Pisa. *SRD'14* consists of 23 subjects, including 16 males and 7 females, aged between 25 and 61 years. The mean age is 45 years and the standard deviation (SD) is 11 years. *SRD'15* consists of 26 subjects, including 14 subjects (11 males and 3 females) from *SRD'14*, aged between 32 and 62 years, with a mean age of 48 years (SD 10 years). The remaining 12 subjects (8 males and 4 females) in *SRD'15*, aged between 29 and 61 years (mean age of 46 years and SD of 9 years), were only recruited for the second campaign. *SRD'15* was collected during system development, with the methods validation based on *SRD'14*, leading to sensors upgrades and changes in the data processing methods. Other publically available datasets were used to further support the Wize Mirror subsystems validation. Information about the datasets used for the validation of each method is given in the corresponding section.

### III. 2D/3D MEASUREMENT FACILITATION

The vast majority of measurements performed by the Wize Mirror are based on data acquired from multiple imaging devices. To facilitate unobtrusive data acquisition and synchronization of the different Wize Mirror sensors, there is

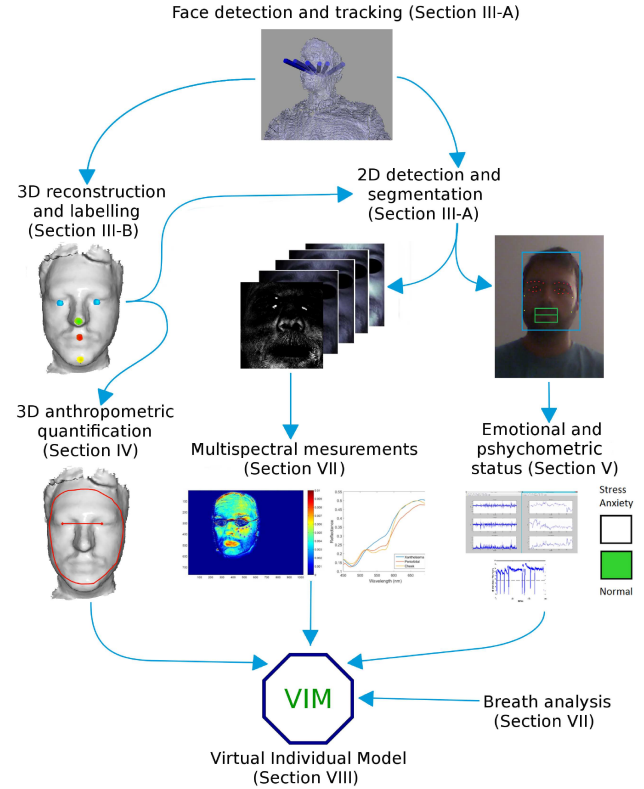


Fig. 3. Multimedia data workflow: from sensors, preprocessing and analysis to virtual individual model calculation.

a need for user face detection, 3D head pose tracking and subsequent face image segmentation (Section III-A). Moreover, to detect and monitor facial changes due to weight, swelling, local growth and facial asymmetry and to perform other bio-morphometric analyses over time, the Wize Mirror is able to perform 3D face reconstruction (Section III-B). Additionally, a 3D face labelling stage is needed to provide different subsequent tasks with the approximate positions of important facial landmarks (Section III-C). A bespoke face recognition system is also implemented to facilitate user access control. A detailed description of the face recognition module can be found in [7].

#### A. Face detection and tracking

Face detection and tracking are performed using depth data from a range sensor as the only input. The proposed face detection and 3D head pose estimation are based on the approach described in [8], the tracking method presented in [9], and a personalized 3D labelled mask explained in Sections III-B and III-C. First, a random forest framework is used to classify depth image patches between two classes (head and no head) and to perform regression in the continuous spaces of the head position and orientation. Then the detection noise is reduced using a Kalman filter method [9]. Finally, the 3D mask is registered to the input range data using an iterative closest point algorithm [10] with the previously estimated pose used for initialization. Mask registration improves the spatial accuracy of the pose estimation and provides the approximate





Fig. 4. Results of the face tracking procedure on a sample of video frames. The 3D labelled mask is projected into 2D images after head pose estimation.

locations of different facial landmarks, such as the eyes, nose, mouth and chin (see the results in Fig. 4). As all the cameras in the system were previously calibrated with respect to the range sensor (as indicated in Section II), the registered 3D mask can be projected into images captured by any of the cameras installed on the Wize Mirror, enabling face segmentation and landmark location in different video streams. This step avoids the redundancy of performing face detection in each video. The computations are reduced because face detection is performed on the depth data stream only, subsequently transforming the vertices of the annotated 3D face mask using rigid transformations and the projective camera model.

### B. 3D face segmentation and reconstruction

As a preprocessing step, a face segmentation method was proposed as the first stage of the reconstruction method. The face segmentation is based on the face pose estimation. With the calculated pose, a 3D model is transformed to match the input depth data. The matched model defines the scan regions which are subsequently used for the 3D face reconstruction.

The implemented reconstruction approach is based on the Kinect fusion method [11]. Originally, the reconstruction method was designed to reconstruct static scenes of rigid objects by moving a range sensor and capturing different points of view of the scene. The reconstruction requirements for the Wize Mirror are different, as the sensor is in a fixed position and the subject is moving. In the proposed algorithm the relative motion of the head is reversed with respect to the sensor to estimate the point of view. This is achieved by using only the output from the face segmentation. More details about the segmentation and reconstruction method can be found in [12].

For the face morphological analysis explained in Section IV the mesh to be processed has to be a manifold, with no holes and no duplicated points or triangles. Although the reconstructed meshes obtained with the mentioned methods are visually good, they, sometimes, do not fully meet all the requirements for a correct morphological analysis (Section IV) of the 3D reconstruction. Therefore, the re-meshing method [13] was applied to the resulting 3D point clouds. Samples of the 3D reconstruction are shown in Fig. 5.

### C. 3D face labelling

Among the previous works on landmark localization reported in the literature, one interesting example is the method described in [14], which uses a point distribution model and performs face detection as the first stage. The model

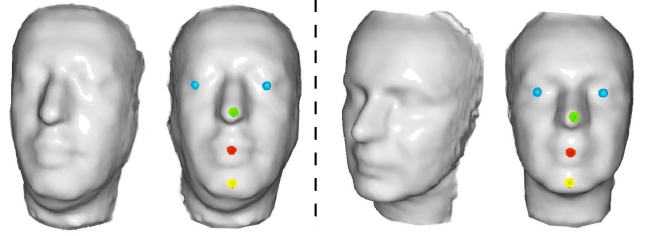


Fig. 5. Results of 3D the face reconstruction and face labelling for two subjects.

presented in [14] did not include landmarks on the mouth or chin. The labelling process for the Wize Mirror produces the approximate positions of the center of the eyes, tip of the nose and centers of the mouth and chin on the reconstructed face. This labelling is an important requirement for morphological analysis, multispectral measurements and analysis of the psychosomatic status as the relevant processing is based on the facial regions defined by these landmarks. The proposed labelling method uses a 3D-deformable annotated model. This model is registered to the reconstructed face as described in [15], which explains how to build the model. The 3D faces are represented by a low-dimensional shape space vector of the statistical shape model, which is calculated through a model-based surface registration process. After the deformable fitting, the labels are near the real landmarks. Then, the closest point to the label from the mesh is selected for each label. The method was tested on *SRD'15*, and all the reconstructions were correctly labelled using the proposed method. Examples of the results are shown in Fig. 5.

## IV. 3D FACE ANTHROPOMETRIC QUANTIFICATION

Anthropometry is the study of body and face morphology. Anthropometric measurements are usually performed manually by trained personnel; therefore, they are often affected by inter- and intra-observer variability. Leslie G. Farkas, a pioneer of modern craniofacial morphology, gathered a set of facial measurements across different ethnic groups, including distances, angles and areas enclosed by anatomical landmarks. Farkas also studied the relations of some syndromes with these measures [16]. Most face morphometry methods are based on 2D images rather than 3D data and require the accurate location of facial landmarks.

One of the challenges of SEMEOTICONS is the automatic computation of geometric descriptors of 3D facial data, acquired via a low-cost scanner, to monitor and quantify an individual's temporal facial shape changes in relation to cardio-metabolic risk and body fat accumulation. There is a relatively large number of techniques reported in the literature based on a sparse set of 2D facial landmarks (often obtained by manual annotation); [17] proposes a method for the prediction of normal and overweight females based on body mass index (BMI). This method uses 2D features (Euclidean distances, angles, and face areas) computed on selected soft-tissue landmarks. The study was extended in [18] by investigating the relation between visceral obesity and facial characteristics to determine the best predictor of normal waist and visceral obesity. Recent



Fig. 6. A visualization of the curves used to define the face measurements implemented in the Wize Mirror. From left to right: MorphoE, MorphoANN, MorphoAB.

technological advances of the depth sensors for 3D acquisition and modeling fostered the employment of digital descriptors from shape analysis to study the morphometric properties on 3D models [19]. Promising results on how to quantify the facial shape variation related to weight gain/loss, are reported in [20], where complex shape descriptors, from the geometric theory of persistent homology, were computed on a subset (23 points) of the Farkas landmarks, and tested on a synthetic dataset of 3D faces.

#### A. Digital measurements

The Wize Mirror is intended to be used by people for the self-assessment of their well-being. Overweight and obesity are among the most relevant factors of cardio-metabolic risk. The requirements of the shape measures are: (i) not requiring the detection of a large number of landmarks (difficult, especially for poorly geometrically characterized landmarks); (ii) being well-defined, easy to implement, and computationally efficient; (iii) being independent of rotation, translation, and scale; (iv) being robust to noise and pose estimation errors. The digital measurements implemented are:

- MorphoE: the length of the maximal curve among those given by the intersection between the face mesh and a family of concentric Euclidean spheres, centered in the nose tip (Fig. 6);
- MorphoG: the geodesic analogue of MorphoE;
- MorphoANN: the area of an annuli at the border of the face mask (likely affected by an increase in subcutaneous fat);
- MorphoAB: the length of a geodesic path in the neck region joining two specific points, under the ears.

MorphoE, MorphoG, and MorphoANN require the detection of only three landmarks (eyes and nose tip), which are automatically located on 3D face meshes (Section III-C). The distance between eyes is used to normalize the measures with respect to the user. MorphoAB requires the location of an additional landmark (chin).

#### B. 3D face measures and physical parameters

Experiments on the reference dataset *SRD'15* established the relation between the described digital measurements and a set of physical parameters related to cardio-metabolic risk.

The subjects had their face reconstructed using both the low-cost depth sensor integrated in the Wize Mirror (see Section

III-B) and a commercial portable structured light scanner (Artec Eva [21]).

The agreement between the reconstructed 3D faces captured by the two scanning platforms was assessed through the intra-class correlation coefficient (ICC) [22]: the ICC values indicate strong agreement for MorphoE, MorphoG and MorphoANN (respectively, .913, .894, and .775), and moderate agreement for MorphoAB (.678).

A set of physical parameters was collected for each subject, according to the literature ([23], [24]): weight, BMI, waist circumference (WC), hip circumference (HC), neck circumference (NC), and fat mass (FM). The proposed digital measurements were computed on facial meshes obtained from both the depth sensor, as described in Section III-B, and the Artec scanner. The Pearson's correlation coefficients between the digital measurements and the physical parameters are reported in Table II for the Artec scans and in Table III for the depth sensor scans. The correlation patterns are similar, indicating that the quality of the depth data does not significantly affect the geometric descriptors.

In both cases, all facial features are highly correlated with weight, BMI and NC (up to  $r = 0.795$ ), and all correlations are highly significant. The correlation with WC and HC is slightly lower but still significant. The correlation for FM is not significant ( $p\text{-value} > 0.1$ ), which may be related to the size and composition of the sample.

TABLE II  
PEARSON'S CORRELATION COEFFICIENTS AND  $p$  VALUES BETWEEN THE PHYSICAL PARAMETERS AND FACIAL FEATURES COMPUTED ON FACE MESHES ACQUIRED USING THE ARTEC SCANNER.

		MorphoE	MorphoG	MorphoANN	MorphoAB
<b>Weight</b>	$r$	<b>.795</b>	.784	.687	.668
	$p$	.000	.000	.000	.000
<b>BMI</b>	$r$	<b>.716</b>	.714	.602	.526
	$p$	.000	.000	.002	.010
<b>WC</b>	$r$	<b>.559</b>	.553	.425	.519
	$p$	.006	.006	.043	.011
<b>HC</b>	$r$	<b>.564</b>	.547	.462	.460
	$p$	.006	.008	.030	.031
<b>NC</b>	$r$	<b>.792</b>	.752	.674	.621
	$p$	.000	.000	.000	.002
<b>FM</b>	$r$	.151	.155	.040	.023
	$p$	.491	.481	.855	.917

#### V. EMOTIONAL STATUS

The Wize Mirror includes methods to analyze facial cues and physiological measures that are related to anxiety, stress (Section V-A) and fatigue (Section V-B).

##### A. Stress and anxiety detection

Stress and anxiety are states of emotional strain that can significantly affect a person's quality of life. According to the literature, there are distinct facial cues that are representative of stress/anxiety that appear in the facial areas of the eyes and mouth, and in the motion pattern of the head [25]. For contactless detection of these cues, a high-resolution camera

TABLE III  
PEARSON'S CORRELATION COEFFICIENTS AND  $p$  VALUES BETWEEN THE PHYSICAL PARAMETERS AND FACIAL FEATURES COMPUTED ON FACE MESHES OBTAINED FROM THE MIRROR 3D SCANNER.

		MorphoE	MorphoG	MorphoANN	MorphoAB
Weight	$r$	.733	.719	.669	.675
	$p$	.000	.000	.001	.001
BMI	$r$	.711	.716	.651	.671
	$p$	.000	.000	.001	.001
WC	$r$	.614	.619	.547	.579
	$p$	.002	.002	.008	.005
HC	$r$	.569	.568	.518	.557
	$p$	.007	.007	.016	.009
NC	$r$	.788	.781	.778	.648
	$p$	.000	.000	.000	.001
FM	$r$	.272	.316	.211	.349
	$p$	.221	.152	.347	.112

is embedded in the Wize Mirror. Advanced video processing algorithms are used to extract and quantify the appropriate facial information to assess a subject's psychophysical status in a reliable and effective manner. The algorithms used for face detection, tracking and region of interest (ROI) segmentation are described in Section III.

The head motion algorithm can detect and quantify the movement of a person's head based on a 2D video file. An ROI, defined as the region of the face between the eyes and mouth, is determined, and the landmarks on the four edges of the ROI are tracked. This ROI was selected as it is characterized by the absence of eyes and mouth movements due to facial expressions; thus, the resulting measurement is only related to head motion. The Kanade Lucas Tomasi (KLT) tracker [26] is applied to track the landmark points, thus creating a time series of the temporal evolution of positions. These time series describe the magnitudes of head motion and velocity, as well as their projections in the horizontal and vertical directions.

Eye-related features are also estimated by the Wize Mirror to detect stressful emotional states [27]. These features, apart from stress, can be modulated by environmental conditions such as temperature and illumination conditions. Active appearance models (AAM) [28] are used to specify landmarks in the eye perimeters that are tracked throughout the video recording. Their distances and relative distribution create a time series that provides eye activity information, including measures of the eye aperture and the rate of blinking, which are known to be correlated with a stressed emotional status [25][27].

Mouth activity is also analyzed in terms of dense optical flow [29] to obtain a description of the motion patterns of the lips. Optical flow is applied only on the Q channel of the YIQ colorspace of the mouth ROI, in which the lips appear brighter than the surrounding tissue [30]. The maximum motion magnitude over the entire video is taken as the source for subsequent feature extraction, including features from the time and the frequency domains [31]. An illustrative example of the analysis framework and the resulting signals is shown in Fig. 7.

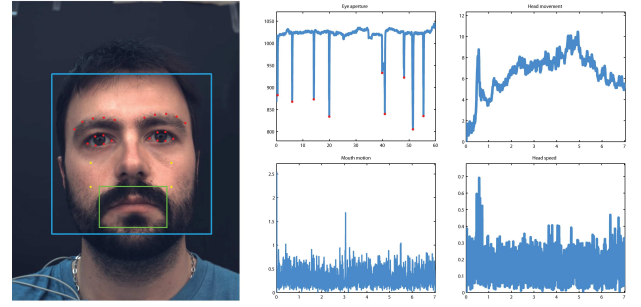


Fig. 7. Left panel: Facial ROI (blue: face, green: mouth) and landmarks (red: eye-related landmarks, yellow: head motion/speed landmarks) determination used for stress and anxiety detection. Right panel: Time series of eye aperture (top left), mouth motion (bottom left), head motion (top right), and head speed (bottom right).

These methodologies and feature estimating algorithms were tested on *SRD'14* and a subset of 24 subjects from *SRD'15*. Each subject performed 12 tasks (3 neutral, 8 stress/anxiety and 1 relaxed states), providing a good initial dataset to evaluate stress/anxiety. Different stressors were used during both acquisition campaigns to investigate various types of stress and anxiety. These features were subsequently used to develop the virtual individual model and to define the personalized wellness index.

### B. Fatigue detection

The Wize Mirror computes a fatigue score that depends on the frequency and duration of yawns, weighted with respect to the time-point of video capture. The yawn detection algorithm is based on real-time tracking of 68 facial landmarks [32]. Yawns are detected by matching landmark-based geometric features of each video frame with templates representing yawning and neutral expressions. Let  $E$  be a stored expression that is represented by a set of templates  $E = (T_1, \dots, T_M)$ .

Then, let  $C$  be the current expression. The similarity score between the expressions  $C$  and  $E$  is the sum of similarity scores between  $C$  and the set of templates of  $E$ :

$$s(C, E) = \sum_{i=1}^M s(C, T_i)$$

where

$$s(C, T_i) = \exp\left(-\frac{\|v(C) - v(T_i)\|_2^2}{\sigma}\right)$$

$v$  is a feature vector encoding landmark positions,  $\|\cdot\|_2$  represents the  $L^2$  norm, and  $\sigma$  is a constant controlling the smoothness of the score distribution. In the experiments reported  $\sigma = 10$  was used. Two vector representations of expressions were implemented:

- A feature vector including the coordinates of all 68 landmarks defined in the Multi-PIE dataset [33]).
- A feature vector encoding information extracted from the mouth region, including: bounding box ratio and distance of the mouth landmarks from the mouth centroid.

The probability that an expression  $C$  represents a yawn is

$$p(C, Y) = \frac{s(C, Y)}{s(C, Y) + s(C, N)}$$



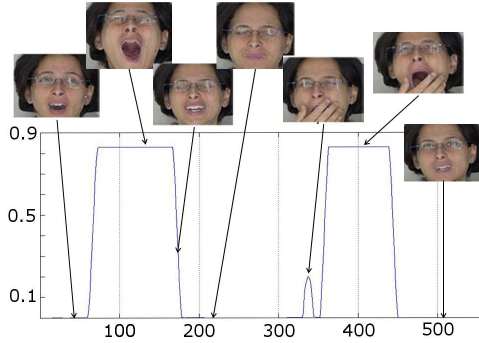


Fig. 8. Yawn probability in a video sequence. The method is robust to occlusion.

where  $Y$  and  $N$  are the stored templates of yawning and neutral expressions, respectively. Only subsequent frames with  $p(C, Y) > 0.5$  for a duration  $> 1$  second are considered a positive detection of yawning. The system uses a mean probability computed on the basis of two vector representations. Fig. 8 shows the probability of an expression representing a yawn for each frame of a video sequence: peaks are correctly located in correspondence to each yawn. The detection is made frame by frame, and the set of subsequent frames belonging to the same yawn are grouped into one *event*. Each event is characterized by its start and duration (number of frames).

The algorithm was tested on two datasets:

1. A set of 10 videos approximately 30 s long acquired in a controlled environment, with a resolution of 1920 x 1080, at 50 fps. The whole set contains 20 *events* (yawns). In this set, 19 events were correctly detected.
2. A set of 10 videos from the YAWDD dataset [34] containing 28 *events*. The videos show tired subjects, simulating driving a car, occasionally yawning. The videos were captured in an uncontrolled environment with natural illumination and challenging subject positions; each video is approximately 1 minute long and was acquired at 30 fps, with resolution 640 x 480. In this set, 21 events out of 28 were correctly detected, 7 events were not detected, and 3 events were incorrectly classified as yawns.

## VI. PHYSIOLOGICAL STATUS

A computational pipeline was constructed to detect and estimate facial semeiotic signs of the user's physiology. The main objective was the development, implementation and optimization of the necessary methods and algorithms to easily and accurately extract heart rate (HR) and respiration rate parameters, as well as to perform heart rate variability (HRV) analysis from color video recordings of the face.

### A. Measurement approach

The proposed procedure exploits the photoplethysmography (PPG) principle [35], a noninvasive optical technique for measuring blood flow. The foundation of this approach is based on the fact that as the heart pumps blood, the volume of blood in the arteries and capillaries varies by a small amount

in synchronization with the cardiac cycle. This variation in blood volume in the arteries and capillaries underneath the skin induce subtle skin color changes. The method explores the use of video recordings of the human face, PPG and ambient or diffused white light to detect the subtle color changes caused by variations in reflected light due to a change in blood volume in the facial region during the cardiac cycle. The initial goal of the color camera-based physiological sign monitoring approach is to estimate the PPG waveform, which is proportional to these changes in skin color, since physiological parameters, such as HR, HRV and breathing rate (BR), can be derived from the acquired PPG waveform. The implemented computational pipeline is presented in [36].

### B. Experimental validation

The computational pipeline was tested on the SEMEOTI-CONS Reference Dataset and the electrocardiogram (ECG) recorded simultaneously with the videos. The results presented in Table IV demonstrate that a highly accurate estimation of the HR from the processing of facial videos is possible. The reported results, applying the computational and analytical pipeline, indicate accuracies on the order of 95-99% compared to the ground truth measurement computed from the ECG signal. The analysis was performed on videos acquired under two recording conditions: a motionless scenario where participants were recorded while mimicking a neutral/resting state and a scenario where participants performed a mental task, in particular the Stroop test [37], which is a scenario that includes natural motion. The method was also tested with participants with deeply pigmented skin both included in the dataset and removed from the dataset. Overall, increased error was observed between trial 8 (4.85% or 5.01%), in which a higher level of motion was observed and trial 3 (1.56% or 1.96%), where the subjects demonstrated the least motion. Moreover, the results indicate that the accuracy of the estimation is significantly influenced, in both recording scenarios, by the inclusion of darker-skinned subjects.

TABLE IV  
THE % ERROR PRODUCED FROM HR MEASUREMENT USING THE SEMEOTI-CONS REFERENCE DATASET.

Recording scenario	Participants	No of recordings	Reference	% error
Mimicking neutral state	Exclude dark skinned	16	ECG	1.56
Mimicking neutral state	Include dark skinned	18	ECG	1.96
Stroop test	Exclude dark skinned	22	ECG	4.85
Stroop test	Include dark skinned	23	ECG	5.01
<b>Total number of recordings</b>		<b>79</b>	<b>Average</b>	<b>3.35</b>

In addition, an evaluation of the HR, respiratory rate and HRV was performed on a subset of 5 participants in *SRD'14* for which five minutes of continuous video recordings were collected. The analysis of HRV was performed using a short recording time (5 minutes) since it is unrealistic to have the subject positioned in front of the mirror for more than a



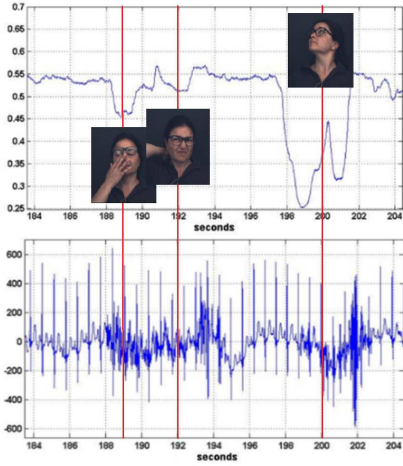


Fig. 9. Motion artifacts on the ECG and video signals due to participant movement (published with permission of the subject).

few minutes. The results confirmed the previously reported outcomes for HR measurements. The HRV analysis was unsatisfactory, with considerable divergence from the values obtained from the ECG analysis (cf. Table V).

TABLE V  
HRV RESULTS (THREE INDICATIVE CASES) FROM 5-MINUTE VIDEO RECORDINGS.

Heart Rate Variability frequency components estimation														
		LF amplitude		HF amplitude		LF power		HF power		LF frequency		HF frequency		
Case	Video	ECG	Video	ECG	Video	ECG	Video	ECG	Video	ECG	Video	ECG	Video	ECG
1	1010	2765	488	705	77	208	56	73	0.08	0.08	0.20	0.24		
2	6288	22463	2475	5060	536	2037	391	853	0.11	0.08	0.22	0.31		
3	6440	4243	1164	395	437	240	209	61	0.04	0.04	0.22	0.34		

These outcomes indicated that although the HR results achieved from the proposed method were satisfactory, implying that the preprocessing phase of the HRV analysis was properly implemented, the constructed signal was not sufficiently representative of the BVP signal or the subsequent processing was not as accurate as required. Finally, measurement of respiration rate was derived from the HF-Frequency band of HRV from both signals (ECG and video) and therefore depended on the accuracy of the HRV estimation.

Visual inspection of the video signals revealed that they were not of sufficient quality for the required signal duration (i.e., five minutes) for accurate HRV analysis. Signal segments were seriously contaminated by noise/artifacts (Fig. 9). Thus, the unreliability of the HRV estimate can be explained in part, by the bad quality and unreliability of the particular recordings, both ECG and video. However, this analysis provided evidence for the feasibility of the proposed method. The reliability should be further tested on datasets designed specifically for HRV analysis, providing ECG and video recordings acquired in parallel for 5 minutes and compatible with basic HRV acquisition requirements.

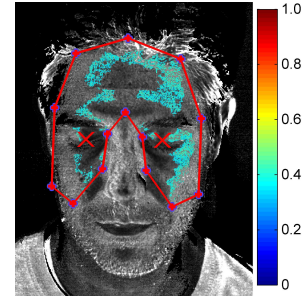


Fig. 10. Concentration of AGE in facial skin (grayscale is a pixel-by-pixel ratio of 475 nm image/ 360 nm image). Colored pixels within the face mask are selected to estimate the AGE concentration.

## VII. MULTISPECTRAL MEASUREMENTS

A multispectral imaging (MSI) system for facial skin analysis has been proposed and evaluated. The MSI system is based on five compact monochrome Flea3 3.2 MP USB 3.0 CMOS cameras (Point Grey) with band-pass filters at selected wavelengths and two computer-controlled LED light sources (white and ultraviolet (UV) light). A heater fan for remote skin heating has also been integrated in the system. The cameras are placed in a three by two pattern adjacent to each other. The camera filters (Edmund Optics) are bandpass filters with center wavelengths and (FWHM) of 360 (45), 475 (50), 560 (10), 580 (10), 650 (50).

Advanced glycation end products (AGE) are linked to inflammation and atherosclerosis and play a role in both the microvascular and macrovascular complications of diabetes. In this project a noninvasive, contactless novel technique is proposed to quantify AGE deposits in skin tissue. The technique collects MSI images during UV exposure from a 365 nm LED (Smart Vision Lights). The method is presented in detail elsewhere [38]. In summary, the AGE level was assessed as the ratio of the fluorescence intensity (475 nm camera) to the illumination intensity (360 nm camera). The image acquisition and processing involved: 1) modulated light sources for ambient light suppression; 2) a ROI mask; 3) avoiding areas with specular reflections (Fig. 10); and 4) a simple calibration procedure. The method was evaluated on data from *SRD'15*. The results from 16 subjects with skin types ranging from fair to deeply pigmented showed that AGE measured using MSI in forearm skin was significantly correlated with the AGE reference method on forearm skin [39] [38]. These results support the use of the technique for contactless measurement of the AGE content in either facial or forearm skin tissue over time.

Lipids accumulate in the skin of persons with increased risk of cardiovascular disease. Xanthelasma is an accumulation of lipids in the periorbital skin. It appears as a soft, yellow lipid skin inclusion. This condition is clearly visible to the naked eye, and can thus in principle be quantified using the MSI system. The skin cholesterol concentration is therefore considered an interesting parameter in the assessment of overall cardio-metabolic risk [40] [41]. The MSI system estimates cholesterol by calculating the cholesterol skin fraction. Detection of cholesterol in the visible spectral range is challenging because

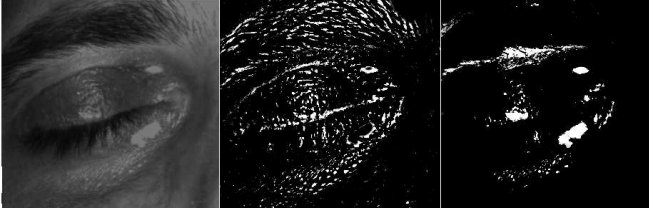


Fig. 11. Cholesterol characterization of subject with high blood cholesterol. Original image (left), small cholesterol deposits enhanced image (mid), and xanthelasma enhanced image (right).

cholesterol does not have a specific spectral signature in this wavelength range. However, it may be detected as a consequence of increased scattering due to the high refractive index in lipids, increased specular reflection and replacement of other skin constituents with cholesterol. The last feature leads to a lower observed blood volume and pigment absorption in lipid-rich areas. The main spectral feature targeted in this project is the scattering change observed due to the high refractive index found in lipids. The proposed method uses a single wavelength (560 nm or 580 nm). The measurement of skin cholesterol is implemented as a cholesterol droplet fraction measure (Fig. 11, mid panel). In this method an area of skin beneath the eye is selected for analysis, and the fraction of pixels covered with spots showing increased skin reflection (white spots) is calculated. The presented algorithm, applied to the data from the same subjects used for the AGE assessment, easily identifies xanthelasma lesions (Fig. 11, right panel).

Endothelial dysfunction is a mechanism that can lead to coronary artery disease. The endothelium balances vasodilation and vasoconstriction during varying blood flow needs. Endothelial dysfunction can result from and/or contribute to several disease processes such as hypertension, hypercholesterolemia, and diabetes. In this project, endothelial function is measured using MSI based on the response of facial skin microcirculation to partial facial skin heating to 39 °C using a controlled heated air flow. Reference recordings of forearm skin during local heating with a fiber optic probe system indicated that the response should be evaluated continuously over the axon reflex period, 1-6 min after the start of heating. The maximum response during this period was used to calculate an index of endothelial function. This index was based on two parameters calculated from the MSI images in the 475-650 nm range: hemoglobin oxygenation ( $SO_2$ ) and the fraction of red blood cells in the skin ( $f_{RBC}$ ; Fig. 12).

The facial skin heating technique, developed using a computer-controlled heated air flow and an IR thermometer to measure facial skin temperature, worked well. This system was tested during the acquisition of *SRD'15*. Normally the temperature at the end of the 10-min period was within 1 °C of the target temperature (range 37.5 – 40 °C). In only a few cases a higher skin temperature was reached for a short period of time, probably due to head movement.

Artifacts were identified due to co-registration difficulties in the MSI setup. This mainly affected the  $SO_2$  images, while the  $f_{RBC}$  displayed less sensitivity. One way to minimize the effect of miss-alignment is to take a spatial average over a

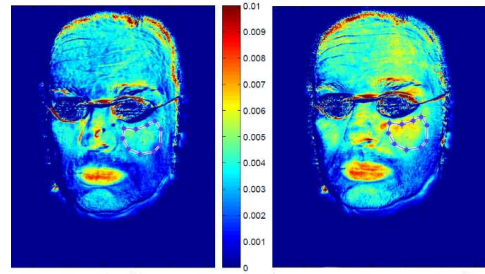


Fig. 12. The fraction of red blood cells in facial skin tissue ( $f_{RBC}$ ; color map indicates a relative scale). The  $f_{RBC}$  is calculated in the selected ROI at baseline (left panel) and 9-10 min after local heating (right panel). A computer-controlled fan focused heated air on the left cheek.

fairly large ROI for each MSI image before calculating the  $SO_2$  and  $f_{RBC}$  estimates. For data from *SRD'15*, all ROI averaged estimates displayed consistent and expected results when exposed to heat.

### VIII. ANALYSIS OF BREATH COMPOSITION

Many studies attempted to find a correlation between breath volatile organic compounds (VOCs) and particular diseases [42], [43], as breath analysis enables monitoring the metabolic processes that occur in the human body in a noninvasive way. This technique is promising and complex; the analysis of the breath gases must take into account a variety of factors: from the subject's posture [44] to the flow rate [45], conditions of the environmental air [46], and the subject's lung volume [47]. For this reason, the majority of studies in this field use sensitive and accurate instrumentation to analyze breath molecules, such as gas chromatography - mass spectrometry [48]. However, such instrumentation is expensive, time-consuming and not easy-to-use.

In the framework of SEMEOTICONS, the challenge was to analyze the user's breath composition by means of a portable, cheap, and easy-to-use device to detect breath gases and analyze them in real-time: the Wize Sniffer [49].

The Wize Sniffer can detect molecules present in the breath that are related to the most noxious habits for cardio-metabolic risk: sensors for carbon monoxide can monitor smoking habits and sensors for ethanol and hydrogen can help the user to maintain their diet under control and avoid an abuse of alcoholic drinks. In addition, variation in the output of oxygen and carbon dioxide gas sensors provides a measure of how much oxygen is retained in the body, and how much carbon dioxide is produced as a by-product of cellular metabolism.

The Wize Sniffer is composed of three modules: one for breath sample acquisition, one for breath molecule detection, and one for data analysis. The acquisition device includes a gas sampling box (600 ml tidal volume [50] and composed of ABS and Delrin) with six gas sensors (manufactured by Figaro Engineering) and widely employed open source controller board (Arduino Mega2560). Since the sensors' output is affected by the water vapor present in the exhaled gas, an HME filter is placed at the beginning of a corrugated tube to reduce the humidity from 90% to 60-70%. The humidity and temperature percentage are also monitored within the sampling



Fig. 13. The Wize Sniffer final configuration. The user blows into the corrugated tube while attempting to keep the expiratory flow rate constant. The flowmeter can be used to calculate the exhaled breath volume. A flushing pump purges the chamber to recover the sensor's steady state between two consecutive measures. The Wize Sniffer was designed to be integrated into the Wize Mirror or to be used as a stand-alone device.

box. Two additional gas sensors with shorter response times work in flowing regime by means of a sampling pump, which operates at 120 ml/s. The sensor outputs are read by the Arduino Mega2560 and sent to the Wize Mirror's main board via an USB connection.

Fig. 13 shows the final configuration and the use of the Wize Sniffer.

The concentration of each of the breath molecules detected by the Wize Sniffer was calculated using a non-linear equation model. The analysis of the gas sensors confirmed their non-linear behavior and their non-selectivity [51], thus impeding the exact discrimination of several substances. Therefore, a more classic approach based on principal component analysis and the K-nearest neighbour classification was adopted. This method uses the sensors' raw data to indicate which "class" the user belongs to, according to their habits ("no risk", "moderate smoker", "heavy smoker", "moderate drinker", "heavy drinker", etc), thus integrating and completing the data derived from the lifestyle questionnaires (see Section IX).

A measurement protocol was drafted for tests on *SRD'15* subjects with different ages, habits, lifestyles and body types. A "mixed respiratory" procedure was selected from available breath sampling procedures [46], [52]. This procedure was selected because the focus was on both endogenous (derived from alveolar exchanges) and exogenous (derived from food and beverages) breath molecules. In addition, since the composition of each breath may vary considerably, to average the breath-by-breath fluctuations in composition, a sampling of multiple (three) breaths was performed.

The KNN classifier correctly classified 20/26 subjects on the basis of their noxious habits. Unfortunately, the classifier was unable to recognize smoking subjects. This may be due to the high detection threshold of the carbon monoxide sensor. This will be solved by replacing the Figaro gas sensor with another sensor with a lower LOD (MQ7 sensor, for example).

While alcohol consumption of up to 1-2 alcohol units/day is not considered noxious (in healthy subjects), smoking is considered very bad for cardio-metabolic risk in all cases.

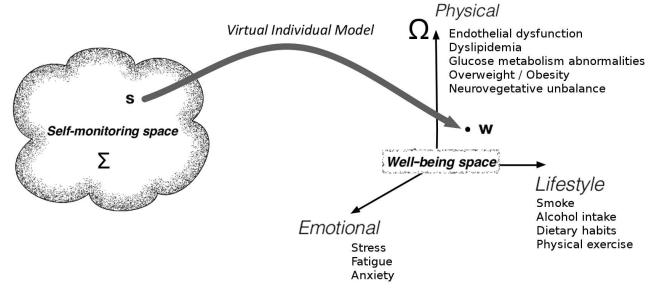


Fig. 14. VIM maps self-monitoring data, represented as vector  $s(t)$ , to vector  $w(t)$  belonging to a 3D well-being space.

## IX. SEMANTIC DATA INTEGRATION AND USER GUIDANCE

The final aim of the Wize Mirror is to monitor individual well-being with respect to cardio-metabolic risk and to foster a healthy lifestyle. According to medical semeiotics of CM risk, the set of computational descriptors previously outlined is expected to convey properly meaningful pieces of information [4]. For completeness, the entire set of descriptors produced by the mirror and the related risk factors are summarized in Table VI. Unfortunately, the direct use of a large heterogeneous set of descriptors is unsuitable for evaluating a user's status with respect to medical knowledge and, even more importantly, to facilitate interaction with the user. This is why the integration process produced by the virtual individual model (VIM) was introduced.

As illustrated in Fig. 14 and detailed in Section IX-A, the VIM maps the data produced by the mirror, which can be considered as point  $s$  in a high-dimensional space  $\Sigma$ , to point  $w$  in semantically structured space  $\Omega$ , called the well-being space. This semantic structure provides the physiological, psychological and behavioral interpretation of the measured parameters. This process partitions the  $\Sigma$  space into three subspaces,  $\Sigma_P$ ,  $\Sigma_E$  and  $\Sigma_L$  named the physical, emotional, and lifestyle-related subspaces, respectively.  $\Sigma_P$  comprises the parameters related to the physical condition of an individual (e.g., blood oxygenation, skin cholesterol, face morphology).  $\Sigma_E$  includes the features (e.g., descriptors of facial expressions, and neurovegetative imbalance) directly related to the emotional status of the subject. Finally,  $\Sigma_L$  is spanned by descriptors on an individual's dietary habits, physical activity, alcohol consumption, and nicotine intake.

According to this partitioning, the well-being space has three axes that describe physical wellness, emotional wellness, and lifestyle-related wellness. To provide the user a concise description of her/his overall condition, the state of the VIM is used to compute a wellness index (WI), which is an appropriately scaled indicator of a user's wellness. It provides the basis for communicating with the user and driving the guidance system.

### A. Virtual Individual Model and Wellness Index

The data collected by the Wize Mirror includes the computational descriptors obtained by mirror sensors and data obtained by user interaction. In Table VI descriptors from the sensors



TABLE VI

THE WIZE MIRROR CARDIO-METABOLIC RISK DESCRIPTORS ALONG WITH THE CORRESPONDING MEASURED PARAMETERS AND LINKED RISK FACTORS. SEE TABLE I TO LINK THE DESCRIPTORS WITH THE SENSORS THAT OBTAIN THE CORRESPONDING DATA.

Descriptor	Measured Parameter	Risk Factor
3DFM	Face geometry	Overweight Obesity
FaDe	Head movement Eyebrow movement Lip movement Yawns Eyelid movement Gaze distribution Blushing Reddening Pallor	Stress, Anxiety Fatigue
PhPa	HR HRV RR	Neurovegetative imbalance
MuMe	Blood volume/oxygenation	Endothelial dysfunction
	Skin perfusion and thermal vasodilation	
	Skin cholesterol accumulation	Dyslipidemia
ExGC	Skin AGE concentration	Glucose metabolism
	Exhaled gas: CO H	Smoking Alcohol intake
	Ethanol	

are listed along with the related data source and the prevalent connected risk factor. In addition, when initializing the mirror, the user is requested to provide a few pieces of data, including: age, weight, height, and waist circumference. The user is also asked to fill out short questionnaires describing their perceived health status and stress conditions (the SF-12 mental and physical components [53] and the Perceived Stress Scale [54]). Finally, questionnaires pertaining to lifestyle habits are also administered to describe dietary habits (DASH [55]), physical activity (IPAQ [56]), alcohol consumption (AUDIT-C [57]), and smoking habits (Fagerström [58]). The latter questionnaires are included in the guidance system (see subsection IX-B for more details).

According to medical knowledge, both computed and user-provided data can be mapped onto a specific axis of  $\Sigma$  space as summarized in Table VII. It follows that VIM implementation is referred back to estimating three mappings,

$$\mathcal{P} : \Sigma_P \mapsto \Omega_P$$

$$\mathcal{E} : \Sigma_E \mapsto \Omega_E$$

$$\mathcal{L} : \Sigma_L \mapsto \Omega_L$$

between the predefined subspaces ( $\Sigma_P$ ,  $\Sigma_E$ , and  $\Sigma_L$ ) of  $\Sigma$  and the three axes ( $\Omega_P$ ,  $\Omega_E$ , and  $\Omega_L$ ) of  $\Omega$ . In this view, estimation of the components of the VIM status is based on the hybrid fusion process depicted in Fig. 15 for the physical axis (analogous schemas were adopted for the other components). The major aim is to correlate VIM status with a medical standard. This led to the adoption of simple but general fusion schema based on the assumption that each component of the VIM status causes changes in the risk factors (see Table VI). For example, the physical component was hypothesized to cause changes in endothelial dysfunction, dyslipidemia, glu-

cose metabolism abnormalities, overweight/obesity and neurovegetative imbalance. These changes, in turn, cause changes in the input. Each of these causal relationships is modelled by a linear equation with the cause(s) as independent variable(s).

Structural equation models (SEMs) were used to implement the association phase [59]. SEMs are widely used in psychometry and behavioral sciences. This choice was motivated by the moderate complexity of a linear SEM, which despite a possible negative influence on the estimation accuracy, is advantageous with respect to overfitting issues. More complex (non-linear) models will be investigated in future developments of the system. Evolution of the data-fusion schema is planned with respect to several aspects, including the use of more general non-linear causal relationships and the adoption of more advanced methods such as those emerging for crowdsourcing and multi-socials integration [60]. These are expected to efficiently cope with partial data inconsistencies, including missing data. Although data incongruence was not observed in the experiments, it can be expected when the mirror is deployed in a general environment.

The complete *SRD* was used to estimate the model coefficients. This dataset includes data from volunteers that are not affected by a known disease. For each subject it contains:

- a complete medical objective examination based on standard clinical testing for CM risk, including four well-established risk scores: Heart SCORE [61], Fatty-Liver index [62], HOMA index [63], FINDRISC index [64];
- nutritional assessment obtained by validated questionnaires and indirect calorimetry;
- psychological evaluation based on validated questionnaires;
- a set of images, pictures, and signals acquired with a suite of sensors operating in experimental conditions close to the Wize Mirror setup that can estimate facial descriptors of CM risk.

This database contains the input data from sensors in the Wize Mirror, user-provided data, and clinical and psychological reference evaluation data, and its usage made it possible to fit the VIM using the SEM maximum-likelihood estimator.

TABLE VII  
CARDIO-METABOLIC RISK FACTOR AND WELL-BEING AXIS.

Risk factor	Well-being axis
Endothelial dysfunction Dyslipidemia Glucose metabolism abnormalities Overweight/Obesity Neurovegetative unbalance Anemia/Plethora/Jaundice	Physical
Stress Fatigue Anxiety	Emotional
Smoking Alcohol intake Dietary habits Physical exercise	Lifestyle

The temporal evolution of the VIM status  $\mathbf{w}$  is expected to provide a concise but complete description of the components of a user's health status. To make it suitable for interaction with general users,  $\mathbf{w}$  is converted into WI using the equation:

$$WI = a_p w_p + a_l w_e + a_e w_l$$



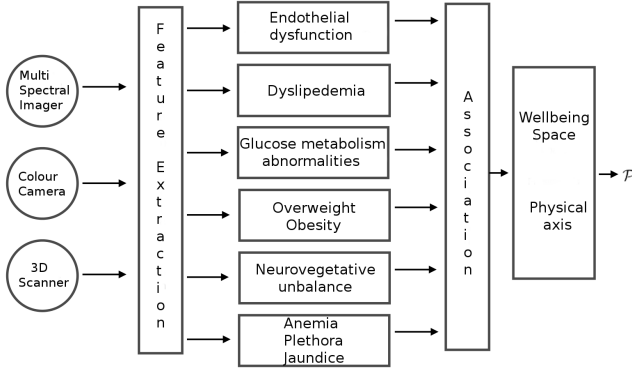


Fig. 15. The schema of data fusion used to estimate the mapping  $\mathcal{P}$  to produce the physical component of the VIM status.

Similar to the SEM method mentioned above, the coefficients  $a_p$ ,  $a_e$ , and  $a_l$  were estimated from the reference dataset by assuming that  $w_p$ ,  $w_e$ , and  $w_l$  depend linearly on  $WI$ .  $a_p w_p$ ,  $a_e w_e$ , and  $a_l w_l$  can be interpreted as physical wellness, emotional wellness, and lifestyle wellness, respectively. The proposed WI is innovative. In contrast to other indices (e.g., WHO-5) based on subjective user evaluation, the proposed WI combines objective measurements (gathered from sensors) with the subjective evaluation of perceived health status performed by the individual. This feature is expected to significantly impact self-monitoring strategies and the effectiveness of user guidance.

To gain an understanding of the relative importance of the two kinds of data (sensed and user-provided) we observe that the estimated values of the three weights are  $a_p = 1.23$ ,  $a_e = 0.50$ , and  $a_l = 0.73$ . By normalizing their sum to unity, the relative impacts on the total  $WI$  of the physical, emotional and lifestyle components are 50%, 20%, 30% respectively. User-provided data are used to set the initial status of the VIM and to compute  $w_l$  that, with the exception of smoke and alcohol consumption which can be measured by the Wize sniffer, can not be captured by the sensors in the mirror. Sensors are therefore responsible for approximately 70% of the  $WI$  computation.

Fig. 16 shows a graphical representation of the WI components as they appear in the Wize Mirror prototype.

### B. Personalized guidance

The Wize Mirror provides customized and personalized suggestions and messages, in accordance with the estimated WI and its change over time, the user profile in terms of attitudes, habits and preferences, and contextual information about the users life circumstances. The guidance has four major lifestyle targets:

- diet,
- physical activity,
- smoking,
- alcohol intake.

Target selection is automatically triggered by the WI. The guidance control includes a set of modulators used for tuning the intensity and the frequency of coaching messages. The

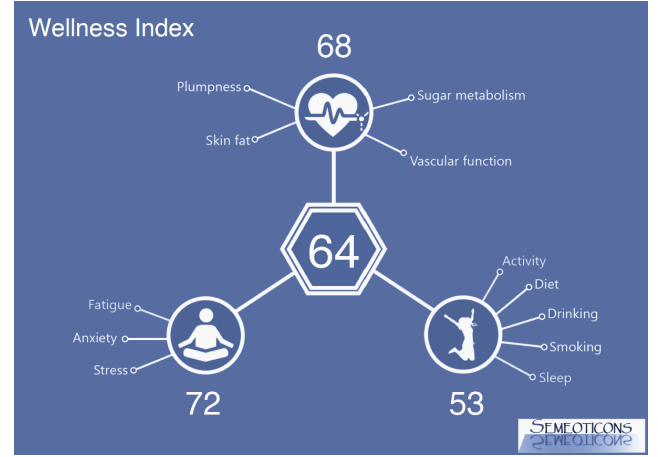


Fig. 16. Graphical representation of the wellness index as it appears in the Wize Mirror prototype. The global wellness (64) is presented along with the physical (68), emotional (72) and lifestyle (53) components.

following questionnaires are administered at the beginning of self-monitoring activity:

- WHO-5 Well-being Index [65] a short questionnaire providing a reliable measure of emotional traits (positive mood, vitality and general interests).
- General Self-Efficacy Scale (GSE) [66] assessment of the perceived self-efficacy and understanding of an individual's motivation to change.

The guidance uses a battery of standardized questionnaires (some previously mentioned) to provide suggestions to promote lifestyle improvement. In particular the Wize Mirror incorporates:

- IPAQ-SF questionnaire [56], [67]: four questions requesting that an individual recall aspects of his/her physical activity over the previous 7 days;
- A revised Alcohol Use Disorders Identification Test Consumption (AUDIT-C) [57]: three questions about alcohol consumption;
- Fagerström Test for Nicotine Dependence [58]: a standard instrument for assessing the intensity of physical addiction to nicotine;
- DASH assessment [55]: a standard instrument for assessing diet;
- Insomnia Severity Index (ISI) [68]: a brief screening measurement of insomnia.

Depending on the score of each behavior a specific suggestion is provided. The messages are tailored to the user's traits to improve the user's communication and engagement. The presentation, visualization and linguistic style of suggestions are tuned in accordance with user's characteristics. To this end a proactive decision support system is under development.

## X. CONCLUSIONS

This paper described the work performed in the European project SEMEOTICONS to develop the prototype of a multisensory platform that detects and monitors facial signs correlated with cardio-metabolic risk, and gives personalized

guidance for lifestyle changes. SEMEOTICONS brings medical semeiotic analysis to everyday life, from the office of medical doctors to the home, the gym and the pharmacy. The empowerment of individuals, in terms of their ability to self-monitor their status and improve their lifestyle, is expected to have a great impact on the reduction of disease burden and health expenditure.

Different technological challenges have been addressed in this work, including contactless analysis of facial signs, multi-modal data integration and development of a virtual individual model. To achieve all the goals, it was necessary to study, implement and test interdisciplinary techniques such as face detection, tracking and reconstruction, morphometric analysis, expression analysis, heart rate estimation, multispectral measurements, cholesterol and AGE estimation. In many cases new approaches were proposed.

The clinical validation of the Wize Mirror is ongoing. The system is being used in a study at three clinical centers involving 60 subjects. This validation study focuses on the reproducibility of measurements provided by the Wize Mirror and the correlation of estimated wellness with respect to cardio-metabolic risk charts (Heart SCORE, Fatty-Liver index, HOMA index, FINDRISC index) which are the groundtruth for the wellness index validation [4].

#### ACKNOWLEDGMENT

This work has been supported by the European Community's Seventh Framework Programme (FP7/2013-2016) under grant agreement number 611516 (SEMEOTICONS).

The authors would like to acknowledge the SEMEOTICONS project partners from Centre de Recherche en Nutrition Humaine Rhone-Alpes, COSMED SRL, DRACO SYSTEMS SL, Hellenic Telecommunications & Telematics Applications Company and INTECS SPA for their valuable contributions to the development of the Wize Mirror.

The authors would like to thank Matija Milanic for performing the measurements during the acquisition campaign and Eleni Kazantzaki and Dimitris Manousos of ICS-FORTH for their assistance in the development and validation of the contactless assessment of anxiety/stress.

#### REFERENCES

- [1] M. Yuan, I. Khan, F. Farbiz, S. Yao, A. Niswar, and M. Foo, "A mixed reality virtual clothes try-on system," *IEEE Transactions on Multimedia*, vol. 15, pp. 1958–1968, 2013.
- [2] A. Rahman, T. Tran, S. Hossain, and A. E. Saddik, "Augmented rendering makeup features in a smart interactive mirror system for decision support in cosmetic products selection," in *IEEE-ACM Symposium on Distributed Simulation and Real-Time Applications*, 2010.
- [3] M. Alhamid, M. Eid, and A. E. Saddik, "A multi-modal intelligent system for biofeedback interactions," in *Medical Measurements and Applications*, 2012.
- [4] G. Coppini, R. Favilla, A. Gastaldelli, S. Colantonio, and P. Marraccini, "Moving medical semeiotics to the digital realm semeioticons approach to face signs of cardiometabolic risk," in *International Conference on Health Informatics*, 2014.
- [5] R. Hartley, "Theory and practice of projective rectification," *International Journal of Computer Vision*, vol. 35, pp. 115–127, 1999.
- [6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1330–1334, 2000.
- [7] S. Klemm, Y. Andreu, P. Henriquez, and B. Matuszewski, "Robust face recognition using key-point descriptors," in *Proc. VISAPP*, 2015.
- [8] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, "Random forests for real time 3d face analysis," *Int. J. Comput. Vision*, vol. 101, no. 3, pp. 437–458, 2013.
- [9] P. Henriquez, O. Higuera, and B. Matuszewski, "Head pose tracking for immersive applications," in *Proc. ICIP*, 2014.
- [10] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 239–256, 1992.
- [11] S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, and R. A. Newcombe, "Kinectfusion: Real-time dense surface mapping and tracking," in *International Symposium on Mixed and Augmented Reality*, 2011.
- [12] Y. Andreu, F. Chiarugi, S. Colantonio, and et. al, "Wize mirror- a smart, multisensory cardio-metabolic risk monitoring system," *Computer Vision and Image Understanding*, vol. 148, pp. 3–22, 2016.
- [13] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Eurographics symposium on geometry processing*, 2006.
- [14] P. Nair and A. Cavallaro, "3-d face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on Multimedia*, vol. 11, 2009.
- [15] W. Quan, B. Matuszewski, and L.-K. Shark, "Improved 3-d facial representation through statistical shape model," in *IEEE International Conference on Image Processing*, 2010, pp. 2433–2436.
- [16] J. Kolar, L. Farkas, and I. Munro, "Craniofacial disproportions in apert's syndrome: an anthropometric study," *Cleft Palate*, vol. 22, no. 4, 1985.
- [17] B. Lee, J. Do, and J. Kim, "A classification method of normal and overweight females based on facial features for automated medical applications," *J Biomed Biotechnol.*, 2012.
- [18] B. Lee and J. Kim, "Predicting visceral obesity based on facial characteristics," *BMC Complementary and Alternative Medicine*, vol. 14, no. 248, 2014.
- [19] E. Vezzetti and F. Marcolin, "3d human face description: landmarks measures and geometrical features," *Image and Vision Computing*, vol. 30, 2012.
- [20] D. Giorgi, M. Pascali, G. Raccichini, S. Colantonio, and O. Salvetti, "Morphological analysis of 3d faces for weight gain assessment," in *Proc. of EG 3DOR 2015, Eurographics 2015 Workshop on 3D Object Retrieval, Zurich (Switzerland) - May 2-3, 2015*, 2015.
- [21] "Artec eva: Fast 3d scanner for professionals." [Online]. Available: <https://www.artec3d.com/3d-scanner/artec-eva>
- [22] C. Nickerson, "A note on a concordance correlation coefficient to evaluate reproducibility," *Biometrics (International Biometric Society)*, vol. 53, no. 4, pp. 1503–1507, 1997.
- [23] S. Millar, I. Perry, and C. Phillips, "Surrogate measures of adiposity and cardiometabolic risk why the uncertainty? a review of recent meta-analytic studies," *J Diabetes Metab*, vol. S11, no. 004, 2013.
- [24] C. Baena, P. Lotufo, M. Fonseca, I. Santos, A. Goulart, and I. Bensenor, "Neck circumference is independently associated with cardiometabolic risk factors: Cross-sectional analysis from elsa-brasil," *Metab Syndr Relat Disord.*, 2016.
- [25] G. Giannakakis, M. Pedititis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. Simos, K. Marias, and M. Tsiknakis, "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89–101, 2017.
- [26] B. D. Lucas, T. Kanade et al., "An iterative image registration technique with an application to stereo vision," in *IJCAI*, vol. 81, no. 1, 1981, pp. 674–679.
- [27] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Computer methods and programs in biomedicine*, vol. 108, no. 3, pp. 1287–1301, 2012.
- [28] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *Pattern Analysis and Machine Intelligence*, 2001.
- [29] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *SCIA*, 2003.
- [30] N. Thejaswi and S. Sengupta, "Lip localization and viseme recognition from video sequences," in *Fourteenth National Conference on Communications*, 2008, pp. 456–460.
- [31] M. Pedititis, G. Giannakakis, F. Chiarugi, D. Manousos, A. Pampouchidou, E. Christinaki, G. Iatraki, E. Kazantzaki, P. Simos, K. Marias et al., "Extraction of facial features as indicators of stress and anxiety," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 3711–3714.

- [32] J. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized mean-shifts," *International Journal of Computer Vision*, 2011.
- [33] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.
- [34] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Harii, "Yawdd: a yawning detection dataset," in *Proceedings of the 5th ACM Multimedia Systems Conference*. ACM, 2014, pp. 24–28.
- [35] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *EMBC*, 2003.
- [36] E. Christinaki, G. Giannakakis, F. Chiarugi, M. Padiaditis, G. Iatraki, D. Manousos, K. Marias, and M. Tsiknakis, "Comparison of blind source separation algorithms for optical heart rate monitoring," in *Wireless Mobile Communication and Healthcare (MobiHealth), 2014 EAI 4th International Conference on*. IEEE, 2014, pp. 339–342.
- [37] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of experimental psychology*, vol. 18, no. 6, p. 643, 1935.
- [38] M. Larsson, R. Favilla, and T. Strömberg, "Assessment of advanced glycated end product accumulation in skin using auto fluorescence multispectral imaging," *Computers in biology and medicine*. [Online]. Available: <http://dx.doi.org/10.1016/j.compbiomed.2016.04.005>
- [39] R. Graaff, R. Meerwaldt, H. Lutgers, R. Baptist, E. de Jong, J. Zijp, T. Links, A. Smit, and G. Rakhorst, "Instrumentation for the measurement of autofluorescence in the human skin," *P Soc Photo-Opt Ins*, vol. 5692, pp. 111–118, 2005.
- [40] Y. Tashakkor and G. B. Mancini, "The relationship between skin cholesterol testing and parameters of cardiovascular risk: a systematic review," *Can J Cardiol*, vol. 29, pp. 1477–87, 2013.
- [41] J. H. Stein, W. S. Tzou, J. M. DeCara, A. T. Hirsch, and E. R. Mohler, "Usefulness of increased skin cholesterol to identify individuals at increased cardiovascular risk (from the predictor of advanced subclinical atherosclerosis study)," *Am J Cardiol*, vol. 101, pp. 986–91, 2008.
- [42] F. Di Francesco, R. Fuoco, M. Trivella, and A. Ceccarini, "Breath analysis: trends in techniques and clinical applications," vol. 79, pp. 405–10, 2005.
- [43] W. Miekisch, J. Schubert, and G. Noeldge-Schomburg, "Diagnostic potential of breath analysis- focus on volatile organic compounds," vol. 347, pp. 25–39, 2004.
- [44] P. Sukul, P. Trefz, S. Kamysek, J. Schubert, and W. Miekisch, "Instant effects of changing body positions on compositions of exhaled breath," 2015.
- [45] F. Di Francesco, C. Loccioni, M. Fioravanti, A. Russo, G. Pioggia, M. Ferro, I. Roehrer, S. Tabucchi, and M. Onor, "Implementation of fowler's method for end-tidal air sampling," 2008.
- [46] B. Thekedar, U. Oeh, W. Szymczak, C. Hoeschen, and H. Partake, "Influences of mixed expiratory sampling parameters on exhaled volatile organic compound concentrations," vol. 5, 2011.
- [47] J. Jones, "The effect of pre-inspiratory lung volumes on the result of the single breath o<sub>2</sub> test," vol. 2, pp. 375–85, 1967.
- [48] D. Guo, D. Zhang, N. Li, L. Zhang, and J. Yang, "A novel breath analysis system based on electronic olfaction," 2010.
- [49] M. D'Acunto, A. Benassi, F. Chiellini, D. Germanese, R. Ishak, M. Margrini, E. Pagliei, P. Pardisi, M. Righi, and O. Salvetti, "Wize sniffer - a new portable device designed for selective olfaction," in *International Conference on Health Informatics*, 2014.
- [50] D. Shier, J. Butler, and R. Lewis, *Hole's human anatomy and physiology*, 11th ed. Mc Graw Hill, 2007.
- [51] P. Clifford and D. Tuma, "Characteristics of semiconductor i. steady state gas response," vol. 3, 1983.
- [52] W. Miekisch, S. Kischkel, A. Sawacki, T. Lieban, M. Mieth, and J. Schubert, "Impact of sampling procedures on the result of breath analysis," 2008.
- [53] The sf-12: An even shorter health survey. [Online]. Available: <http://www.sf-36.org/tools/sf12.shtml>
- [54] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of Health and Social Behavior*, vol. 24, pp. 385–396, 1983.
- [55] National Heart, Lung, and Blood Institute. Description of the DASH eating plan. [Online]. Available: <http://www.nhlbi.nih.gov/health/health-topics/topics/dash/>
- [56] A. Mannocci, V. Bontempi, V. Colamesta *et al.*, "IPAQ-SF: Reliability of the telephone-administered international physical activity questionnaire in an italian pilot sample," *Epidemiol Biostat Pub Health*, vol. 11, pp. e8860–1, 2014.
- [57] Public Health England Alcohol Learning Resources. [Online]. Available: <http://www.alcohollearningcentre.org.uk/News/NewsItem/?cid=6150>
- [58] C. Pomerleau, S. Carton, M. Lutzke, K. Flessland, and P. OF, "Reliability of the fagerstrom tolerance questionnaire and the fagerstrom test for nicotine dependence," *Addict Behav*, vol. 19, pp. 33–9, 1994.
- [59] D. W. Kaplan, *Structural Equation Modeling Foundations and Extensions (Advanced Quantitative Techniques in the Social Sciences)*. SAGE Publications, 2000.
- [60] X. Song, L. Nie, L. Zhang, M. Akbari, and T.-S. Chua, "Multiple social network learning and its application in volunteerism tendency prediction." New York, NY, USA: ACM, 2015, pp. 213–222.
- [61] J. Perk, G. e. a. De Backer *et al.*, "European guidelines on cardiovascular disease prevention in clinical practice (version 2012). the fifth joint task force of the european society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts)," *Eur Heart J*, vol. 33, pp. 1635–1701, 2012.
- [62] G. Bedogni, S. Bellentani *et al.*, "The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population," *BMC Gastroenterol*, vol. 6, p. 33, 2006.
- [63] D. Matthews, J. Hosker, A. Rudenski, B. Naylor, D. Treacher, and R. Turner, "Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man," *Diabetologia*, vol. 7, pp. 402–9, 1985.
- [64] J. Lindström, M. Peltonen, J. Eriksson *et al.*, "Determinants for the effectiveness of lifestyle intervention in the finnish diabetes prevention study," *Diabetes Care*, vol. 31, pp. 857–62, 2008.
- [65] The WHO-5 website. WHO-5 Well-being Index. [Online]. Available: <https://www.psykiatri-regionh.dk/who-5/who-5-questionnaires/>
- [66] U. Scholz, B. Doña, S. Sud, and R. Schwarzer, "Is general self-efficacy a universal construct? psychometric findings from 25 countries," *European Journal of Psychological Assessment*, vol. 18, pp. 242–251, 2002.
- [67] L. Crinière, C. Lhommet, A. Caille, B. Giraudeau, P. Lecomte, C. Couet, J. Oppert, and D. Jacobi, "Reproducibility and validity of the french version of the long international physical activity questionnaire in patients with type 2 diabetes," *J Phys Act Health*, vol. 8, pp. 858–65, 2011.
- [68] C. Bastien, A. Vallières, and C. Morin, "Validation of the insomnia severity index as an outcome measure for insomnia research," *Sleep Med*, vol. 2, pp. 297–307, 2001.