A Holistic Approach to Interpreting Human States in Smart Environments Providing High Quality of Life

Adrian Stoica, Yumi Iwashita, Chris Assad, Michael S. Ryoo
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91107
Adrian.stoica@jpl.nasa.gov

Gareth Howells
School of Engineering and Digital Arts
University of Kent
Canterbury, UK
W.G.J.Howells@kent.ac.uk

***Abstract*— We formulate a concept of a future smart environment for high quality of life (SEQUAL) that would empower humans to compensate for physical and cognitive disabilities associated with sickness and aging. In SEQUAL the assessment of the state of 'well-being' - from behaviors and biological signals - is holistic, meaning that the estimation of individual's health, emotional condition, activity and wishes, are from the beginning determined in relation to each other and in (individual's own) context, with superior results compared to when estimated independent from each other, as in common practice. Similarly, the prediction of a person's future condition, intentions, future needs, and actions/treatment/interventions are determined holistically. SEQUAL includes robots, mobility systems and assistive devices for physical intervention, as well as remote professional caregivers, family and friends, to provide intelligent assistance and support network, aiming for higher quality of life for both patient and caregiver.**

Keywords—human condition; quality of life; human bio-signals; assistive technologies

I. INTRODUCTION

An important application of future smart environments is to monitor, interpret and enhance quality of life. In general terms it includes the collection of information from humans, its 'understanding' in the context of supporting their needs, and providing appropriate actions. Such topics are treated in the context of Smart Environments (SE), Ambient Intelligence (AmI), Ambient Assisted Living, etc. SE and AmI are fast growing mutually complementary areas with a huge potential to benefit society – the SmE is a place enriched with technology (sensors, processors, actuators, information terminals, and other devices interconnected through a network) for "Smart Homes" but also hospitals, cars, etc.; AmI enhances the global behavior of such a system by providing high level functionality which provides an added value to the typical services expected in a specific environment [1]. Ambient intelligence (Aarts and Marzano, 2003) is characterized by being: (a) *Embedded, (b) Context aware*: recognizes people/context (c) *Personalized*: tailored to individual needs (d) *Adaptive:* can change in response to individual's needs (e) *Anticipatory*: can anticipate desires. Finally, AAL technologies aim specifically to supporting elderly and disabled people in their environment, with affordable, easy to use and meaningful ICT tools [2]. The context of this paper overlaps these areas, and it is extended as we include in the discussion robots, mobility systems and assistive devices for physical intervention. Remote professional caregivers, family and friends, to provide intelligent assistance and support network are also included.

Expressions of human states could be unintentional (e.g. activities, emotions), or intentional (e.g. verbal commands and hand/body gestures). Current systems are built around the individual recognition of these independently of each other, largely as a consequence of being specialized equipment from manufacturers with different orientations and focus. While a global perspective can be obtained by fusion at various levels, there is a difference between fusing end products and using various modalities in all stages of processing. One should also recognize it it is relevant to combine the observation with prior knowledge, to achieve improved recognition, prediction, determine appropriate action.

This paper proposes a concept of a future smart environment for high quality of life in which the assessment of the state of 'well-being' is holistic from the start, meaning that the estimation of individual's health, emotional condition, activity and wishes, are from the beginning determined in relation to each other and in (individual) context, with superior results compared to when estimated independent from each other.

The need for holistic view has been previously identified, e.g. in [3] which rremarks that existing approaches to AAL often fail to consider a human agent's needs from a holistic perspective – tthey propose a framework based on a model for Activity-centered modeling of knlwiedge and interaction tailored to users (ACKTUS).

The paper is organized as follows. Section 2 introduces Smart Environment for high QUAlity of Life (SEQUAL). Section 3 introduces a unifying formalism bringing together information independently analyzed by different fields: biometrics, health monitoring, activity recognition, etc. The holistic approach facilitates optimal resource allocation, sharing and inter-

modality cuing for obtaining the necessary information to 'understand' the human state. In Section 4 we propose the idea of enhancing the amount of human-centered information by using, in addition to the direct signal, indirectly derived signals (e.g. projections/reflections), traditionally treated as noise and hence eliminated. We demonstrate an advantage in the context of a classification based on gait, showing that using information from (body) shadow dynamics in additional to that from body dynamics leads to an improved correct classification rate. The effect of using the shadow is equivalent to that of using body information obtained from a second sensor (camera). Section 5 discusses the use of robots, mobility systems and assistive devices for physical intervention, and of remote 'users' (professional caregiverrs, family and friends), to provide intelligent assistance and support network, aiming for higher quality of life for both patient and carers, and associated issues of privacy and security

II. SEQUAL

Future SEQUAL seek to provide a fundamental role in increasing people's quality of life, empowering them to exceed current physical and information processing limitations, while also optimizing resources. This power comes from high capabilities in information processing, global communication and accessibility of specialized medical databases, which will allow future smart environments to be connected to global cyber-physical systems and to extract more value from data originating in sensors distributed in what surrounds us, what is on us (wearable), and inside us (as implants). These advances come as a natural extension of technology developments in the last decades. SEQUAL components, sensors, processors would be distributed in the walls, objects, in wearable devices, and in implants. The sensing system will include rich modalities of sensing signals from human bodies, including new generations of imaging sensors in many bands, chemical sensors, metabolic sensors, etc.

The greatest remaining challenge is to integrate and make sense of all the data in order to take the best action; more computational power does not automatically bring 'understanding' or 'cognition'. Core to enabling quality human-oriented services from future SEQUAL will be the ability to correctly interpret human states and commands, predict their consequences, intentions and needs.

Understanding will be followed by action/intervention and this combination of perception-reasoning-action will be in effect continuous. Robots would provide a mobile sensing capability, i.e. accessibility to areas that lack built-in sensors but also to provide additional support or help to humans. Built-in automation (from temperature control to control of wheelchair or exoskeleton) will also be part of the intervening actions. The system would also allow remote involvement of stakeholders: remote professional caregiverrs, family and friends, to provide intelligent assistance and support network.
A subset of information sources of human state, is shown in Fig 1.

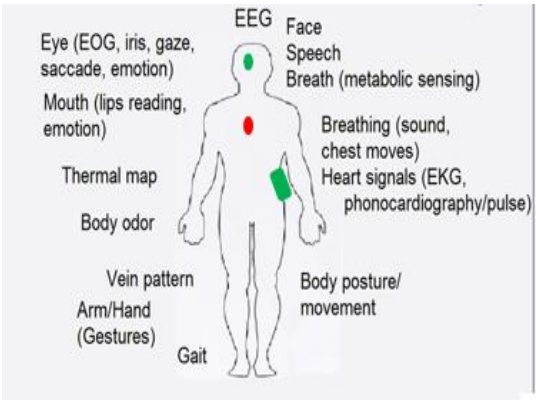

Fig. 1 Various sources of information, seamlessly captured by future SEQUAL/smart environments, offer complex information about the individual.

Fig. 2 A pictorial representation of elements of SEQUAL: seamless and on/in patient-sensors, robots, remote operators and viewers, a cloud of resources

III. A UNIFYING FRAMEWORK OF HUMAN STATE

A. ***A holistic perspective***

Human-oriented analysis is currently disparate over a number of research areas. Biometrics is mostly focused on determining identity: who the individual is. Activity recognition is concerned with determining what the individual is doing. Medical/healthcare/psychometric systems focus on determining the health/emotional condition of the individual. Human-computer/robot interfaces/interaction focus primarily on recognizing the requests/commands (communication) from the human. Speech offers a clear example of diversity of objectives for analysis. From the spoken words "Come here!" one could attempt to identify the speaker, qualify his health (voice intensity, tremor), and emotion/mood (intonation, pitch) and the semantics/meaning of his communication (what is his request/command).

*The human state signal (**S**) is modeled as a multi-dimensional time-varying vector*, composed from (and potentially decomposable in) a superposition of information vectors (components), including: identity characteristics (**I**) , health

characteristics (**H**), emotional state (**E**) and message(activity), **(A)**[1], and other factors (**O**). *Each in turn is a superposition of other vectors*, e.g. a person may be speaking and gesturing at the same time, health may be affected by more than one condition, etc. The components are written in increasing order of variability: identity is least changed with time, health changes at shorter intervals of time, emotion even shorter, etc.

$$\mathbf{S = I + H + E + A + O}$$

As an example, when a speaker says 'Listen to me' the sound contains a richness of information and can be used for multiple analysis for different objectives. Example of outcomes: Biometrics/Speaker ID module: John. Health/state estimator: (voice intensity, tremor): Healthy, tired. Emotion/mood estimator (intonation, pitch) : excited. Semantics/meaning estimator: a request to listen.

As traditionally only one of the components is targeted by analysis (e.g. **I** is targeted by biometrics identification), processing focuses in isolating a single component of **S**. (so other components are considered noise from this perspective and filtered out).
Spatial and temporal filters are usually applied as a first step, with parameters around signal location in space, and tuned for the specific time-scale.

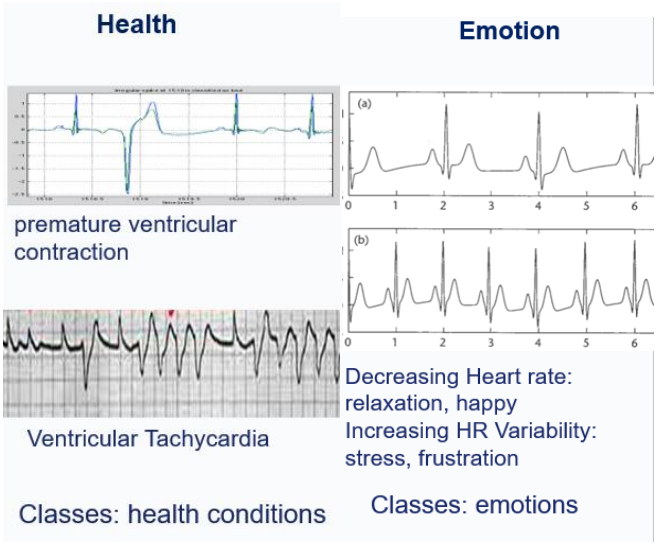

Fig. 3 Signal interpretation depends on the objective of the analysis – illustration for ID, health and emotions

We propose to perform multiple human-oriented analysis approaches under a unifying framework. We believe more information can be used and also the computation in signal processing has a lot of algorithmic communality and same dedicated processors can be used.

For example, the majority of subsequent processing steps are similar, whether the target is identification or activity recognition or another, but with different parameters, features, databases of templates, etc. In the case of biometrics one determines and compares with a database of individuals/identities; in the case of health one determines and compares with databases of symptoms; and for a spoken message we compare with a database/repository of words. This is relevant because it allows use of similar processing architectures, and it influences information fusion mechanisms.

The next section analyzes two of the components that come from the same source of information, i.e. body movement, but are used for two different objectives: determining identity I and activity **A.** The study can be extended to other modalities.

B. Observing human body, from biometrics and activity recognition perspectives

Biometrics identification is concerned with determining the identity of an observed person, for whom specific characteristics that differentiate him from others are known. Human activity recognition is an automated understanding of body-part movements of humans [4]. In both cases, the data being processed are videos obtained from cameras (single or multiple). In biometrics the output is an identity associated with an individual; in activity recognition the outputs are activity labels corresponding to the motion performed by the actors. While activity recognition extends to human-human interactions, and even group activities, here we refer to recognition of single person gestures and actions.

Gait is one of the biometric modalities based on human body, and it can be performed from a distance. Gait recognition approaches generally fall into two main categories: (1) model-based analysis and (2) appearance based analysis. Model-based approaches use parameters of gait dynamics, such as stride length, cadence, and joint angles, which are detected using background subtraction and body-part matching [9]. Appearance-based approaches use measurements of gait features from silhouettes obtained by feature extraction methods, such as gait energy images [10].

There have been two major directions in recognizing human activities. The first is to estimate joint locations of human body-parts (e.g. limbs) per image frame, and model human activities as a sequence of such body-part parameter (e.g. joint angle) changes [5]. A human posture is detected for each frame using background subtraction and body-part matching, and their sequence was represented with probabilistic models (e.g. hidden Markov models). The second category corresponds to the approaches that treat video observations as high-dimensional data and apply machine learning methodologies to analyze them [6,7,8], without explicit understanding of human body part status. One can again decompose these approaches

into two types, approaches using global features [6] and those extracting local spatio-temporal features [7,8].

C. Commonality in algorithmic processing

In general, processing of human activity recognition algorithms can be described as follows. Initially, given a raw sequence of image frames (i.e. videos), the system considers it as an observation and uses it directly for the recognition. Segmentation (e.g. foreground subtraction or body-part region estimation) methodologies are applied depending on the approaches. Next, features are extracted by capturing local and global motion in video data. Particularly, approaches using spatio-temporal features do not require any segmentation processing and are directly extracted from videos. Generative as well as discriminative classifiers have been widely tested. Table 2 summarizes processing steps of activity recognition and gait-based identification, emphasizing the similarities

Table 1 Evaluation of identity [I] and activity recognition [A]. Source of information is entire body; camera sensors (video/infrared, single or stereo).

Approaches	Representation	Algorithm/Function	Performance	Challenges
Model-based	3D body model (e.g. stick figure)	Statistical sequential modeling (e.g. [5]), k-nearest neighbor [G1]	[I] 80~90% side-view walking [A] 90~95% for hand gestures and simple actions	[I] occlusion, changes of appearance changes due to clothes change [A] complex activities with multiple actors
Global appearance-based	Template (Global features)	Template matching [6], discriminative classifiers [G]	[I] 95~% for 100 people database [A] more than 90% for simple actions with static background	[I] appearance changes due to clothes change and walking direction change [A] Moving backgrounds
Local appearance-based	Bag-of-words (local features)	Generative (e.g. pLSA [8]) and discriminative (e.g. SVMs [7] [G]) classifiers	[I] 95~% for 100 people database [A] 90~100% for simple actions (e.g. running); ~70% for complex multi-person activities	[I] appearance changes due to clothes change and walking direction change [A] View-point invariance

Table 2 Similarity in sequence of processing algorithms for gait-based identification [I] and activity recognition [A]

Processing step\Component	Gait-based identification [I] and activity recognition [R]	Obs.
Segmentation	Foreground regions are estimated from videos	[I] Some of model-based approaches do not need any segmentation [R] Recent spatio-temporal features [7,8] do not require any segmentation.
Extract feature	[I] Features capturing motion and body shape are extracted [R] Features capturing salient motion of persons are extracted	[I] In general appearance-based approach shows better performance than model-based approach. [R] 3-D body model representation [5] enables view invariant extraction of human features, but obtaining it is difficult.
Classification	[I] Person labels are obtained [R] Activity labels are obtained	Machine learning approaches

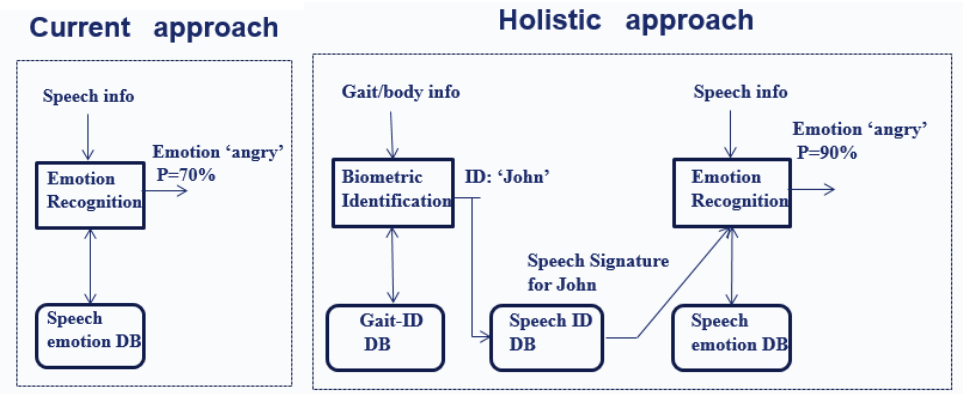

Fig 4. Increasing recognition rates based on information from other components/modalities.

D. *Increasing recognition rates based on information from other components/modalities : Inter-Modality Cueing*

Knowledge of one component of the S vector helps in identification of other components. For example, if we identify a specific emotion from gait analysis, we could compensate for it - to increase the recognition of identity from gait biometrics, but also to increase speaker recognition by filtering out emotion in speech signal. Conversely, if we know who the individual is we could increase the recognition of his/her emotion, and the recognition of the words he speaks. Assume one detects emotion from speech – comparing the features extracted from speech with features in a speech emotion database, and classifies the person as being angry, with a probability of 70%. In the holistic approach, if gait information is also available one can perform gait identification and determine the specific individual, and then use the information about the specific individual (whose voice may be in fact quite mellow in general), and deduce that in fact he is angry with higher probability, say 90%. (Fig. 4)

Here observation of individual's gait leads to an identification; his speech record (template/characteristics) are pulled from the database and provided to the emotion recognition engine, together with actual speech fragment, from the same individual and for whom emotion is to be detected. Compared to a simple recognition using the speech emotion database, which is averaged/integrated over all speakers (shown on the right) the case when identity is known allows to better characterize his emotion

IV. CONCLUSIONS

We proposed a holistic view on interpreting human state and suggested how it can be used to increase the quality and efficiency of the analyses involved in this interpretation.

V. ACKNOWLEDGEMENT

The work undertaken in this paper was facilitated by the Royal Society under their International Exchanges Scheme

REFERENCES

[1] Journal of Ambient Intelligence and Smart Environments – IOS Press

[2] The New Everyday: Views on Ambient Intelligence Paperback – February, 2003

[3] D. Surie et al, Agent-Supported Assessment for Adaptive and Personalized Ambient Assisted Living Trends in Practical Applications of Agents and Multiagent Systems. Volume 90 of the series Advances in Intelligent and Soft Computing pp 25-32

[4] J. K. Aggarwal and M. S. Ryoo, "Human Activity Analysis: A Review", ACM Computing Surveys (CSUR), 43(3), April 2011.

[5] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian Process Dynamical Models for Human Motion", IEEE T PAMI, 30(2), February 2008.

[6] A. F. Bobick, J. W. Davis, "The Recognition of Human Movement Using Temporal Templates", IEEE T PAMI, 23(3), March 2001.

[7] I. Laptev, "On space-time interest points", Int Journal of Computer Vision, 64(2/3), 2005.

[8] J. C. Niebles, H. Wang and L. Fei-Fei, "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words ", Int Journal of Computer Vision, 79(3), 2008.

[9] D. Cunado, M. Nixon & J. Carter: Automatic Extraction and Description of Human Gait Models for Recognition Purposes, CVIU, Vol.90, No. 1, pp.1-41, 2003.

[10] J. Han & B. Bhanu: Individual recognition using gait energy image, Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, No. 2, pp.316-322, 2006.

[11] Yumi Iwashita, Adrian Stoica, Ryo Kurazume, Person Identification using Shadow Analysis,British Machine Vision Conference, pp.35.1--10, September, 2010

[12] S. Dağtaş, G. Pekhteryev, Z. Şahinoğlu, H. Çam and N. Challa, Real-Time and Secure Wireless Health Monitoring Int J Telemed Appl. 2008; 2008: 135808.

[13] C Assad, MT Wolf, A Stoica, T Theodoridis, K Glette *"BioSleeve: A Natural EMG-Based Interface for HRI"*, Proc ACM/IEEE International Conference on Human Robot Interaction, Mar 3-6, Tokyo, 2013

[14] A .Kokosy, T. Floquet, G. Howells, H. Hu, , M. Pepper, M. Sakel,, C. Donze **"SYSIASS – an intelligent powered wheelchair"**First International Conference on Systems and computer Science, Villneuve d'Asq, Lille, France 2012

[15] M. Gillham, G. Howells **A dynamic localized adjustable force field method (DLAFF) for real-time assistive non-holonomic mobile robotics** International Journal of Advanced Robotic Systems, **12**:147 2015 October, doi 10.5772/61190