Probabilistic Multisensory Emotion Estimation Framework for Assistive Robotic Applications

Dario L. Sancho-Pradel[1,2]
[1]Surrey Space Centre, University of Surrey
Guildford, Surrey, GU2 7XH, UK
d.sancho@surrey.ac.uk

Elena G. Nikolova[2]
[2]Cognitive Neuroinformatics, Bremen University
Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
nikolova@informatik.uni-bremen.de

Abstract

Computer-based emotion recognition is an emerging field with envisioned applications ranging from customer satisfaction evaluation to human-machine interaction. In this paper we present a general framework for continuous emotion inference based on Bayesian biometric data fusion and the circumplex model of affect. We apply this framework to the field of assistive robotics focused on elderly and impaired people who require a wheelchair for mobility purposes. The objective is to provide an emotion-based safety layer that complements the classical collision avoidance approaches typically included in these systems. In many real-case applications the calculation of the emotional valence is not feasible, and therefore we also present here a promising novel context-based alternative currently under development.

Introduction

Humans seem most capable (probably due to self introspection) of distinguishing certain emotional states in other individuals just by glancing at them for a short period of time, extracting behavioural, facial and other relevant cues. Computers on the other hand, lack such capabilities and therefore need to be trained for that purpose. Due to the difference between the perceptual and analytical abilities of humans and computer, however, cues that are most suitable for us (e.g. gestures, poses, expressions) may result significantly less useful for a computer system. On the other hand, some other information of a more "numerical nature" (e.g. electrocardiogram measurements), not that indicative for humans, could provide a computer system with key insights into the emotional state of an individual.

The work we present here is the result of our preliminary research into continuous emotion estimation and its application to the development of an emotion-based navigation assistive layer for an intelligent wheelchair typically operated by a physically and/or cognitively impaired elderly person (our target user). This layer is completely user-centric and provides a technology that the user utilises unconsciously. Its task is to adapt the level of support that the user receives from the wheelchair based on the physical and cognitive capabilities of the patient, his/her driving performance, the navigation context and the user's biometric readings.

Our work differs from those of other authors in the field in a number of aspects. Firstly, it is concerned with real-time estimation of emotions, and as such, it presents techniques and mathematical models that can operate online. Secondly, while most of the related works are concerned with user satisfaction and digital entertainment, our goal is to develop robotic systems capable of reacting and adapting their behaviour in real-time (e.g. modify the navigation strategy of a mobility platform) based on the user emotions. Thirdly, due to the nature of our application and target population, we are restricted in the type and number of sensors we can use. Finally, we propose the use of context-based information in order to solve the problem of valence calculation without using intrusive or uncomfortable sensors.

Motivation

Electrical wheelchairs, scooters and walkers are probably the most widespread modern assistive devices used by elderly and disabled people in order to increase their mobility. Although collision avoidance has been recently included in smart wheelchairs (Minguez and Montano 2004), (Mandel, Huebner, and Vierhuff 2005), (Simpson 2005), tests with the target population show a clear need for plasticity in the platform's safety layer and navigation module (Annicchiarico et al. 2007). As a result, researchers are looking into shared control and shared autonomy paradigms (Barrue, Cortes, and Annicchiarico 2007). The objective is to dynamically adapt the assistance provided by the platform to the precise level of support required by the user. During our initial studies, we realised that technology anxiety and difficulty to interact with computer-based systems were common in our target population and could prevent an individual from using or fully exploiting a clearly beneficial technology. For instance, a person driving through a door may perform well if the corridor is empty since he/she feels relaxed. On the other hand, a busy corridor makes him/her feel anxious and the cognitive and motor actions required to drive the wheelchair through the same door may become too demanding. Moreover, stress and anxiety can make the user unable to give any meaningful command through the platform's graphical user interface (GUI). In this case, the emotion estimation algorithm presented here would notice that the user needs assistance and would, for instance, (depending on the user's medical condition) present a very simplified

GUI with just a couple of clear options, or would grant full control to the autonomous navigation system. In general, the estimated emotional level of the user would be passed to the shared control or share autonomy module and fused with other data in order to decide how much autonomy is granted to the navigation system.

Emotions and Biometrics

Experiments in affective neuroscience research have yield two main theories or models for understanding emotions and their related disorders. The traditionally dominant one is the theory of basic emotions, which posits that humans are evolutionary endowed with a discrete and limited set of basic emotions (Ekman 1992), (Panksepp 1998), (Tomkins 1962 1963). According to this theory, each emotion is independent of the others in its behavioural, psychological and physiological manifestations, and each arises from activation within unique neural pathways of the central nervous system (Posner, Russel, and Peterson 2005). One of the main criticisms towards this theory comes from its failure to explain the near ubiquitous comorbidity among mood disorders (Posner, Russel, and Peterson 2005), and the fact that usually emotions are not recognised as isolated discrete mental states but as combinations of overlapping ones without tangible discrete borders (Russell and Fehr 1994). The circumplex model of affect, on the other hand, proposes a two-dimensional approach where each emotion can be understood as a linear combination of two fundamental neurophysiological systems, namely arousal and valence (Russell 1980). Arousal is related to alertness, and can be seen as a measure of activation, while valence can be understood as a measure of how positive or negative an individual feels about something. In that way, emotional states could be represented as a Cartesian product of these two parameters. As Figure 1 shows, the two parameters cannot be considered fully independent from each other (particularly in the extremes of the axis). However, as an often valid approximation, arousal and valence levels are typically calculated separately, obtaining in that way quantitative estimates of the emotional level of an individual. The techniques and methods presented in this paper focus on the characterisation of the arousal and valence levels of an individual, based on the circumflex model of affect.

Emotions have been highly correlated in the psychophysiological literature with facial muscle activity and sympathetic nervous system activity (Bradley 2000), (Lang et al. 1993), and therefore they could be estimated, in principle, from biometric measurements. The most common biometric indicators used in the literature are shown next.

- Heart Rate (HR) and Heart Rate Variability (HRV). This information can be extracted from electrocardiogram (ECG) readings and pulse oximeters. HR and HRV are highly correlated with arousal, and HRV can be used as an indicator of mental load, long term stress and heart wearing. Blood Pressure (BP) is typically linked with arousal and stress.
- Galvanic Skin Response (GSR) is a popular method of measuring the electrical resistance of the skin, which is highly correlated with arousal.
- Respiratory frequency measured as chest expansion and contraction, has also been related to arousal.
- Skin Temperature (ST). Stress can modify the blood distribution in the body (e.g. extremities), which can result in local temperature variations.
- Facial muscle activity extracted from Electromyogram (EMG) readings. Facial expressions extracted from cameras.
- Speech cues extracted from microphones.
- Eye gaze and movement, extracted either from cameras or glasses, are typically used in attention assessment experiments.

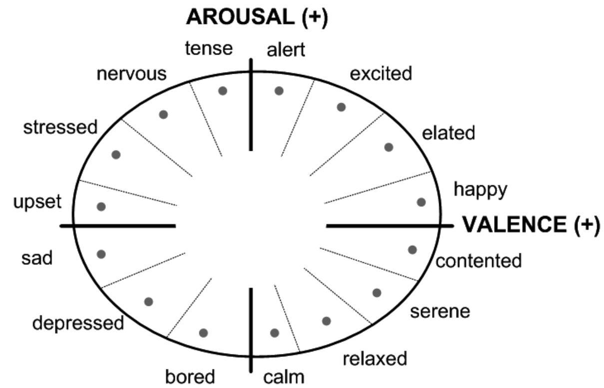


Figure 1: A graphical representation of the circumplex model of affect, based on (Russell 1980).

When emotion estimation is used in psychological experiments or in applications such as user satisfaction analysis (Fragopanagos and Taylor 2005), (Desmet and Hekkert 2007), (Mandryk and Atkins 2006) the choice of sensors is less constrained since the user will wear them for a relatively short period of time, probably during a finite set of sessions. On the other hand, continuous monitoring of signals in an individual places clear restrictions in terms of comfort, non-invasiveness and reliability, and therefore, certain clinical devices such as blood pressure devices, EMGs or academic ones such as eye tracking glasses, cannot be considered. Other sensors such as video cameras, could also be excluded due to privacy issues. The environment sets also constraints on the sensors, since for instance noisy and crowded environments would probably be inappropriate for performing reliable speech analysis. Finally, the gender and age of the user could definitely influence the choice of sensors. Due to the nature of our application, we concentrate on biometric measurements of the ”numerical type” such as HR, HRV, GSR, Breathing or ST, complemented with the contextual information that the robotic platform can provide us with. Human-machine interaction cues can provide a measurement of its emotional, physiological and cognitive state. For instance, a person driving an electric wheelchair will often have a joystick as input device. Useful information can

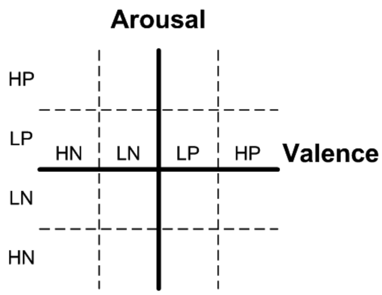


Figure 2: Example of a simplified discretisation of arousal and valence. H=high, L=low, P=positive, N=negative.

be extracted from simple frequency analysis of the joystick's commands (e.g. tremor or lack of coordination), and a more elaborated context-based processing can reveal, utilise and even quantify more complex factors such as visual neglect, stress due to a imminent complex task (e.g. narrow corridor navigation), and the location of walls and obstacles.

Probabilistic Emotion Estimation Based on Bayesian Inference

As we have seen, emotions are directly related to arousal and valence levels, two quantities that unfortunately cannot be directly measured. Instead, we infer them from the readings of a set of biometric sensors and contextual information. Here we present a novel probabilistic recursive algorithm for estimating emotions in a continuous fashion that we call the Probabilistic Emotion Estimation Filter. This algorithm makes no assumptions about the type or number of sensors. It only requires that each sensor is capable of roughly relating its measurement history with arousal and/or valence levels. In order to simplify the notation and since the following analysis is applicable to both arousal and valence, we will refer to them indistinctively as the *state* (x) to be estimated, where $x \in X = \{$ *arousal*, *valence* $\}$. Furthermore, we discretise the state, so only a finite set of different values are considered. This is consistent with the fact that we try to differentiate between a finite set of emotions. We consider the set of state values $\xi = \{\chi_1, \chi_2, ..., \chi_N\}$,and represent the state at time t as a state vector of N components, i.e. $x_t = \{x_{d,t}\}$ for all $d = 1, 2, ..., N$. Figure 2 illustrates this discretisation for the case of $N = 4$ and $\xi = \{HN, LN, LP, HP\}$. Now we are in a position to reformulate the estimation of emotions as the calculation of the probability vector

$$P(x_t|s^1_{1:t}, ..., s^n_{1:t}) \equiv \{P(x_{d,t}|s^1_{1:t}, ..., s^n_{1:t})\} \tag{1}$$

where $d = 1, 2, ..., N$ and $s^i_{1:t}$ is the complete measurement history of the $i-th$ sensor, at time t. In order to simplify the notation further, we make the following definitions. Note that the following vectors are column vectors.

Belief state vector. It represents the system's internal knowledge about the state based on the updated history of the sensors' measurements.

$$\begin{aligned} bel(x_t) &= P(x_t|s^1_{1:t}, ..., s^n_{1:t}) \\ &\equiv \{P(x_{d,t}|s^1_{1:t}, ..., s^n_{1:t})\} \end{aligned} \tag{2}$$

where $d = 1, 2, ..., N$. If only the i-th sensor is considered in the state estimation then we use the notation

$$bel(x^i_t) = P(x_t|s^i_{1:t}) \tag{3}$$

Prediction state vector. If only past measurements were take into account, the state could not be estimated but predicted

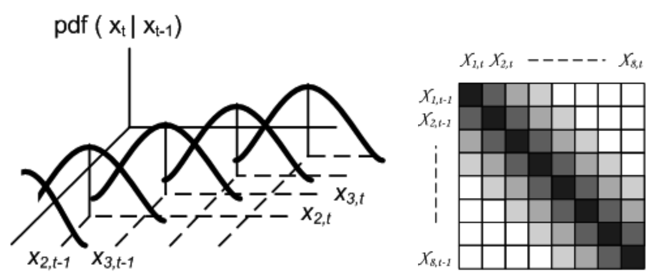
$$\begin{aligned} \overline{bel}(x_t) &= P(x_t|s^1_{1:t-1}, ..., s^n_{1:t-1}) \\ &\equiv \{P(x_{d,t}|s^1_{1:t-1}, ..., s^n_{1:t-1})\} \end{aligned} \tag{4}$$

for all $d = 1, 2, ..., N$. Similarly, if only the i-th sensor is considered in the state prediction then we use the notation

$$\overline{bel}(x^i_t) = P(x_t|s^i_{1:t-1}) \tag{5}$$

State transition probability matrix $P(x_t|x_{t-1}) = M = [m_{i,j}]$. M is an $N \times N$ square matrix, whose elements $m_{i,j}$ represent the probability of reaching the state i from state j. In this paper we assume that M is stationary, i.e. $M \neq f(t)$. A normally distributed M is depicted in Figure 3.

$$P(x_t|x_{t-1}) = M = \begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,N} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & \cdots & \cdots & m_{N,N} \end{pmatrix} \tag{6}$$



(a) 3D view of a normal M

(b) 2D projection for a normal 8x8 M.

Figure 3: Graphical representation of the state transition matrix M where the transitions are assumed to follow a normal distribution. In figure (b) darker areas represent higher transition probabilities. Note that the different Gaussian curves need not to have the same parameters.

If we want a continuous state estimation, we need to be able to update the state with every new peace of sensor evidence. For that purpose, we perform a Bayesian recursive state estimation. Next we briefly state three useful probabilistic identities.

The Bayes rule of conditional probability for two conditions

$$P(a|b,c) = \frac{P(b|a,c)P(a|c)}{P(b|c)} = \frac{P(c|a,b)P(a|b)}{P(c|b)} \tag{7}$$

By the law of total probability

$$P(a|b) = \sum_i P(a|x_i, b)P(x_i|b) \tag{8}$$

The identity for conditional independent random variables a and b

$$P(a, b|c) \quad = \quad P(a|c)P(b|c) \tag{9}$$

As mentioned before, our objective is to estimate arousal and valence levels, and we can do so by calculating the probability vector $P(x_t|s^1_{1:t}, ..., s^n_{1:t})$. We first derive the solution for the case of $n = 2$, i.e. $P(x_t|s^1_{1:t}, s^2_{1:t})$, which will be then generalise for N sensors. In the following section, unless stated explicitly, the multiplication operation between belief vectors is done on a per component basis, and do not refer to dot products. First we notice that expanding the conditional part of $P(x_t|s^1_{1:t}, s^2_{1:t})$ and applying (7) we obtain

$$\begin{aligned} &P(x_t|s^1_{1:t}, s^2_{1:t}) = \\ &P(x_t|s^1_t, s^2_t, s^1_{1:t-1}, s^2_{1:t-1}) = \\ &\frac{P(s^1_t, s^2_t|x_t, s^1_{1:t-1}, s^2_{1:t-1})P(x_t|s^1_{1:t-1}, s^2_{1:t-1})}{P(s^1_t, s^2_t|s^1_{1:t-1}, s^2_{1:t-1})} \end{aligned} \tag{10}$$

Assuming conditional independence between the measurements of different sensors

$$\begin{aligned} &P(s^1_t, s^2_t|x_t, s^1_{1:t-1}, s^2_{1:t-1}) = \\ &= P(s^1_t|x_t, s^1_{1:t-1})P(s^2_t|x_t, s^2_{1:t-1}) \end{aligned} \tag{11}$$

so (10) can be rewritten as

$$\begin{aligned} &P(x_t|s^1_{1:t}, s^2_{1:t}) = \\ &\frac{P(s^1_t|x_t, s^1_{1:t-1})P(s^2_t|x_t, s^2_{1:t-1})P(x_t|s^1_{1:t-1}, s^2_{1:t-1})}{P(s^1_t, s^2_t|s^1_{1:t-1}, s^2_{1:t-1})} \end{aligned} \tag{12}$$

At this point, our route to calculate the belief distribution $P(x_t|s^1_{1:t}, s^2_{1:t})$ differs from the traditional measurement update step based on $P(s^i_t|x_t)$, commonly found in probabilistic robotics (Thrun, Burgard, and Fox 2005). The main reason is the lack of a $P(s^i_t|x_t)$ model available relating biometric sensors and emotions. Trying to obtain one would require medical trials that would extend beyond the time frame of this research. It is also very possible that the validity of such model would be confined to a particular individual and that due to comorbidity and changing medical conditions, such model would most likely require a very complex state vector definition and higher order Markovian relations. Also, it is worth noticing that most biometric measurements, such as heart rate or GSR, lack the level of uncertainty inherent to the distance measurements performed by ultrasonic or laser range sensors.

We propose instead, to work with $P(x_t|s^i_{1:t-1})$ and a state transition matrix M. The first one requires each sensor to estimate the state based on its own measurement history. Since we have a set of past measurements available, we can use a variety of approaches, such as pattern recognition techniques or fuzzy inference, in order to obtain and continuously update this posterior model.

Applying the Bayes rule to $P(s^1_t|x_t, s^1_{1:t-1})$ and $P(s^2_t|x_t, s^2_{1:t-1})$, and gathering terms, (12) becomes

$$\begin{aligned} &P(x_t|s^1_{1:t}, s^2_{1:t}) = \\ &\frac{P(s^1_t|s^1_{1:t-1})P(s^2_t|s^2_{1:t-1})}{P(s^1_t, s^2_t|s^1_{1:t-1}, s^2_{1:t-1})} \cdot \\ &\frac{P(x_t|s^1_{1:t})P(x_t|s^2_{1:t})P(x_t|s^1_{1:t-1}, s^2_{1:t-1})}{P(x_t|s^1_{1:t-1})P(x_t|s^2_{1:t-1})} \end{aligned} \tag{13}$$

Since the law of total probability must be obeyed, we can gather together all the terms that are not dependent on x_t in a normalisation constant η, obtaining

$$\begin{aligned} P(x_t|s^1_{1:t}, s^2_{1:t}) \quad &= \quad \eta \frac{P(x_t|s^1_{1:t})P(x_t|s^2_{1:t})}{P(x_t|s^1_{1:t-1})P(x_t|s^2_{1:t-1})} \cdot \\ &\cdot \quad P(x_t|s^1_{1:t-1}, s^2_{1:t-1}) \end{aligned}$$

, which can be rewritten in belief terms as

$$P(x_t|s^1_{1:t}, s^2_{1:t}) \equiv bel(x_t) = \eta \cdot \frac{bel(x^1_t) \cdot bel(x^2_t)}{\overline{bel}(x^1_t) \cdot \overline{bel}(x^2_t)} \cdot \overline{bel}(x_t) \tag{14}$$

for each component of x_t. This result can be generalised as follows

$$P(x_t|s^1_{1:t}, ..., s^n_{1:t}) \equiv bel(x_t) = \eta \cdot \prod_{i=1}^{N} \left(\frac{bel(x^i_t)}{\overline{bel}(x^i_t)} \right) \cdot \overline{bel}(x_t) \tag{15}$$

for each component of x_t. The normalisation constant η can be calculated applying the law of total probability

$$\sum_{x_t \in \xi} bel(x_t) = 1$$

thus obtaining

$$\eta = \left(\sum_{x_t \in \xi} \left(\prod_{i=1}^{N} \left(\frac{bel(x^i_t)}{\overline{bel}(x^i_t)} \right) \cdot \overline{bel}(x_t) \right) \right)^{-1} \tag{16}$$

Now we need to find the expressions for the different beliefs involved in (14). Applying the law of total probability to $\overline{bel}(x_t)$ and assuming the state transition to be a first-order Markov process

$$\begin{aligned} &\overline{bel}(x_t) \equiv P(x_t|s^1_{1:t-1}, s^2_{1:t-1}) = \\ &= \quad \sum_{d=1}^{N} P(x_t|x_{d,t-1})P(x_{d,t-1}|s^1_{1:t-1}, s^2_{1:t-1}) \end{aligned} \tag{17}$$

for each component of the vector $\overline{bel}(x_t)$. Since $P(x_t|x_{d,t-1})$ represents the $d-th$ column of matrix M (i.e. the state transition probability vector from the $d - th$ state to any other state), we finally obtain

$$\overline{bel}(x_t) \equiv P(x_t|s^1_{1:t-1}, s^2_{1:t-1}) = M \cdot bel(x_{t-1}) \tag{18}$$

where here $\cdot$ represents the matrix-vector product. Operating in a similar manner we obtain

$$\begin{aligned} \overline{bel}(x^i_t) \quad &\equiv \quad P(x_t|s^i_{1:t-1}) \\ &= \quad \sum_{d=1}^{N} P(x_t|x_{d,t-1})P(x_{d,t-1}|s^i_{1:t-1}) \qquad (19) \\ &= \quad M \cdot bel(x^i_{t-1}) \qquad (20) \end{aligned}$$

Finally, Table 1 presents the recursive algorithm that implements the Probabilistic Emotion Estimation Filter for an arbitrary number of biometric sensors.

Emotion_Estimation_Filter$[bel(x_{t-1}), bel(x^i_{t-1})]$

1. $\overline{bel}(x_t) = M \cdot bel(x_{t-1})$
2. **for all** i **do**
3. Obtain $bel(x^i_t)$
4. $\overline{bel}(x^i_t) = M \cdot bel(x^i_{t-1})$
5. **endfor**
6. $ubel(x_t) = \prod_{i=1}^{N} \left(\frac{bel(x^i_t)}{\overline{bel}(x^i_t)} \right) \cdot \overline{bel}(x_t)$
7. $\eta = \left(\sum_{x_t \in \xi} ubel(x_t) \right)^{-1}$
8. $bel(x_t) = \eta \cdot ubel(x_t)$

Return $[bel(x_t)]$

Table 1: Recursive implementation of the Probabilistic Emotion Estimation Filter.

Calculation of $bel(x^i_t)$

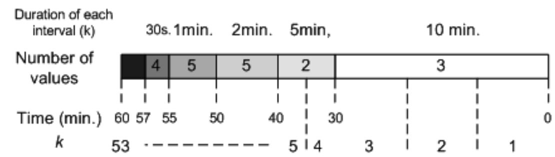
The posterior probability of each of the components of $bel(x^i_t) \equiv P(x_t|s^i_{1:t})$ represents the probability of the state vector based only on the updated set of measurements of the $i-th$ sensor. In other words, what *line 3* of Table 1 requests from each sensor is its belief about the current state vector for arousal or valence. In general, one can safely assume that newer measurements provide much more information than older ones, and therefore, we could make the following approximation

$$bel(x^i_t) \equiv P(x_t|s^i_{1:t}) \simeq P(x_t|s^i_{t-n:t}) \quad (21)$$

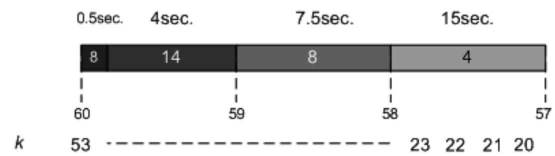
For reasonable values of n, this probability can be empirically approximated using suitable training data collected from a particular individual. We could increase n several orders of magnitude without incurring into much bigger computational load, by noticing that since old measurements are less relevant in the estimation of the state than recent ones, it would be reasonable to provide measurement averages rather than single values. Therefore, dividing the interval $[t-n:t]$ into k subintervals ($k \ll n$), we could approximate $bel(x^i_t)$ by

$$bel(x^i_t) \equiv P(x_t|s^i_{1:t}) \simeq P(x_t|\rho^i_{1:k}) \quad (22)$$

For instance, if the first interval contains r elements then $\rho^i_1 = average(s^i_{1:r})$, where $r \leq n$. Notice, that in general the number of measurements averaged on each interval could be proportional to how far in time their measurements are, hence increasing the resolution around recent measurements. Figure 4 presents an example of this multiresolution interval decomposition. This approach presents the advantage that $P(x_t|\rho^i_{1:k})$ can be learned for each individual subject using for instance time series prediction architectures such as ANFIS (Jang 1993). For a detailed description of different time series prediction techniques and relevant machine learning algorithms see for instance (Chatfield 2003) and (Witten and Frank 2005). As a final comment, we want to note that although some sensors cannot infer by themselves arousal or valence (e.g. the room temperature sensor j cannot infer $bel(x^j_t)$), their readings may clearly influence the estimation $bel(x^i_t)$ of sensor i (e.g. the GSR sensor i). In this cases, sensor j should not be included in the filter, but should be included in the calculation of $bel(x^j_t)$.



(a) General Decomposition



(b) Interval decomposition for the last 3 minutes

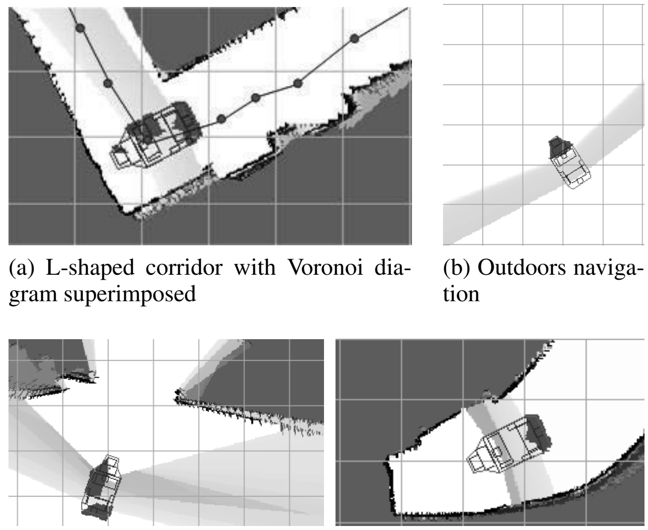
Figure 4: Approximation of $bel(x^i_t)$ based on a multiresolution interval decomposition of the measurements taken in the last 60 mins. Clearly, old values influence less, as shows the fact that for the first 30 min. of the interval, only three averages of 10 mins. each are considered (k_1 to k_3). If we get for instance 10 samples per second, then $n = 60 \cdot 60 \cdot 10 = 36000$, while $k = 53$.

Valence Calculation for Wheelchair Control Applications

Evidently, the calculation of $bel(x^i_t)$ depends on what are we trying to estimate and the type of the sensor we are using. A number of approaches relating arousal with biometric measurements such as HR, respiration or GSR, have been proposed cf. (McQuiggan, Lee, and Lester 2006), (Anttonen and Surakka 2005), (Herbelin et al. 2004), (Mandryk and Atkins 2006), and therefore we will not elaborate further on it. There is however, one difference worth mentioning between these works and our approach. It is well known that HR measurements are very much influenced by the level of physical activity of the individual. Although for obvious reasons this effect is significantly reduced in the case of wheelchair driving, our calculation of $bel(x^{HR}_t)$ takes into account arm movements by means of an accelerometer located on the wrist.

The estimation of valence, is altogether a completely different matter. Inferring the subjective consideration of something being pleasant or not based on a set of biometric measurements has proved to be a very complex task. Its calculation cannot be neglected since as Figure 1 shows, similar arousal levels are indicator of contradictory emotional states such as happy or upset, elated or stressed, excited or nervous. Most of the works found in the literature related to valence estimation are based on either questionnaires, speech processing cf. (Jones and Jonsson 2005), (Cichosz and Slot 2005), (Vidrascu and Devillers 2007), or facial expression extraction from EMG sensors cf. (Mandryk and Atkins 2006). Clearly, none of these approaches are valid in the case-scenario of wheelchair control by elderly and/or impaired users, as the valence has to be estimated on-line, the patient is not continuously (or at all) communicating verbally with the wheelchair, and EMG sensors are certainly intrusive for daily use. Since under these conditions we are in no position to obtain reasonable values for valence, we propose a modified version of the circumplex model of affect that uses what we call a *Context Based Valence* (CBV). The CBV can be seen as a simplified version of the valence, whose value is based on the context of the task being performed. In our case, the task is driving an electrical wheelchair to perform activities of daily living and it is subdivided into a set of subtasks such as negotiating an obstacle, navigating through a door, following some signs, etc. The context is the environment surrounding the wheelchair, its configuration and obstacles, the medical history of the user (e.g. is he/she cognitively capable of localising him/herself outdoors) and his/her current driving performance. We propose to extract this information (except the medical history) from the range sensors that are typically mounted on autonomous and semiautonomous wheelchairs. Our experimental wheelchair Rolland (Mandel, Huebner, and Vierhuff 2005) has two laser scanners, one on the front and one on back, which are used for collision avoidance, semi- and autonomous navigation. Furthermore, wall segmentation is performed online, based on the raw measurements of these sensors. The current framework can also compute Voronoi diagrams either from local sensor-based grid maps or from a pre-existing global grid map derived from a CAD blueprint. Doors of a predefined size can be also detected based on the segmented data and the Voronoi diagram (Mandel, Huebner, and Vierhuff 2005). We are currently working on extending this framework to enable the recognition of contextual information that could help estimating the valence of the driver. Figure 5 shows local evidence grids obtained in real-time while driving Rolland, which clearly show contextual information. Figure 5a for instance presents a relatively narrow L-shaped corridor, whose sharp turn may be cognitively challenging to navigate for many elderly people with certain impairments. Moreover, even though the security layer of Rolland will drive fairly straight the corridor (even if the driver tries to collide against a wall), there are psychological factors (e.g. claustrophobia) that could still arise negative emotion in the user.

By means of a performance function based on the number of collisions avoided by the wheelchair, joystick commands, time spent, etc., we can categorise a set of environments that the user finds particularly challenging. For instance, when a person that has had difficult times driving through a narrow L-shaped corridor is confronted with a similar configuration, a neutral valence will be initially inferred. However, as soon as his arousal levels increase over certain level we will assume a negative valence. The context of Figure 5b, i.e. outdoors navigation, could be somehow opposite. Many elderly with impairments find the short time they spend outdoors enjoyable, and therefore an initial positive valence is assigned to this context, even for positive arousal levels. The medical history of the user is part of the contextual information. If the driver has, for instance, localisation problems at cognitive level, or suffers from agoraphobia, an outdoor experience could be very challenging, and therefore high arousal levels would assign negative values to the valence. Another example is presented in Figure 5c, where the driver is approaching an open door. In this context, the challenging part is to pass through the door and therefore, high arousal levels in the approaching phase and poor performance function will be associated with low valence. On the other hand, after the maneuver is finished, the valence will be reset to neutral and only very high levels of arousal will force a reconsideration. It is important to understand that the nature, shape and values associated with each context have to be learned



(a) L-shaped corridor with Voronoi diagram superimposed

(b) Outdoors navigation

(c) Approaching an open door

(d) Dead-end corridor

Figure 5: Four different contexts for valence inference, based on the evidence grids generated online by Rolland.

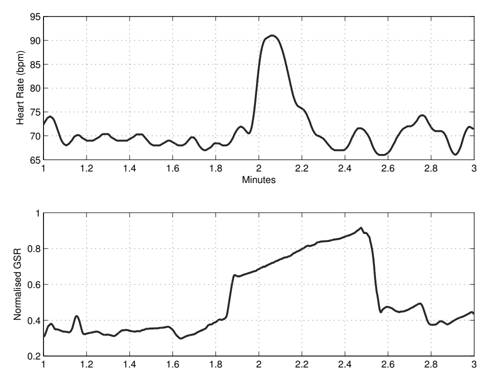
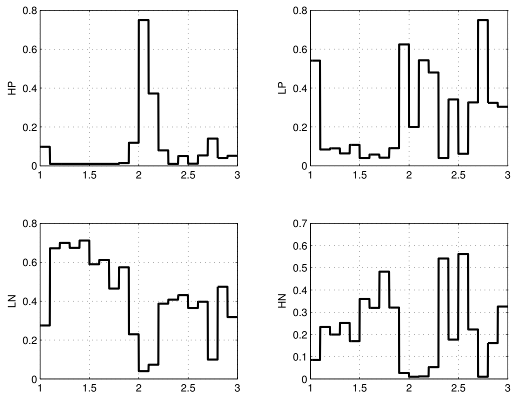
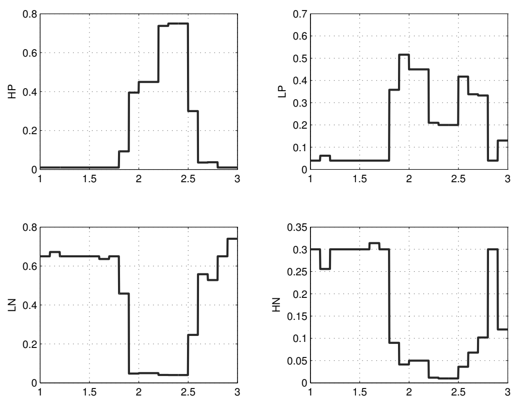


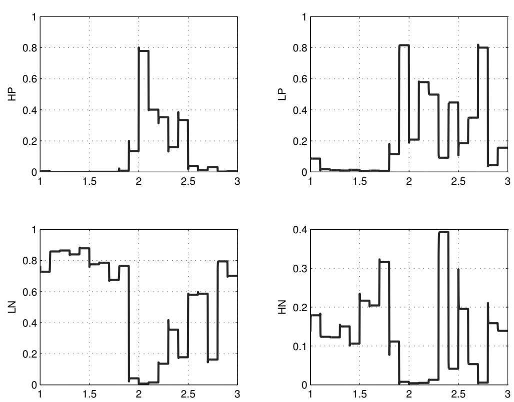
Figure 6: Biometric recordings (HR and GSR) during an experiment where the user encountered unexpectedly the end of the corridor as shown in Figure 5d .



(a) $\{bel(x = \text{arousal}_t^{HR})\}$



(b) $\{bel(x = \text{arousal}_t^{GSR})\}$



(c) Inferred $\{bel(x = \text{arousal}_t)\}$ using the algorithm presented in Table 1

Figure 7: Results of the probabilistic arousal estimation based on the biometrics from the dead-end corridor scenario. In this analysis the arousal has been discretised as shown in Figure 2. (a) and (b) are rough estimations obtain via simplified fuzzy inference. (c) Despite that and the coarse discretisation of the arousal, the algorithm successfully localises in time the stress event.

from experience for each particular user. Furthermore, if a map of the area navigated is available, sections that systematically present a challenge to the user can be marked and CBV estimations could be done automatically.

Preliminary Results

In order to illustrate the application of the algorithm shown in Table 1 to the recursive estimation of arousal, and the idea of context-based valence extraction, we present an extract of an experiment where a non-impaired subject is driving Rolland. At certain point while driving fast, the subject encountered an unexpected dead-end in a corridor (see Figure 5d). In Figure 6 we show the HR and GSR measurements during a three minutes interval around the dead-end corridor event (shortly before minute 2). Figure 7a and 7b present individual sensor beliefs, which are used by the recursive probabilistic emotion filter for inferring the driver's arousal (Figure 7c). The individual sensors beliefs $\{bel(x = \text{arousal}_t^{HR})\}$, $\{bel(x = \text{arousal}_t^{GSR})\}$ used in this example corresponds to a very simplified fuzzy inference models where only the latest measurement is used. Despite the rough estimation of arousal provided by the inference system, the coarsely discretized ($N = 4$) state vector ξ and the use of the generic state transition matrix M shown in Figure 3b, the algorithm clearly localises in time the stress point of the user. A fine tuning of the transition matrix for each user would produce a smoother filter response. The CBV approach makes sure that similar emotion responses to dead-end corridors would flag such locations in a way that negative valence would be inferred in case of high arousal levels.

Conclusions and Future Work

This paper presented a novel probabilistic framework for multisensory emotion estimation, introduced the context-based valence calculation, and illustrated its application in the field of assistive robotics. The preliminary results with non-impaired subjects were promising, even with coarse state discretisation and transition matrices. The next step is to gather experimental data from the target population (impaired elderly people) and use it to develop algorithms for identifying biometric particularities of individual subjects, and applying them to refine $bel(x_t^i)$. Context recognition, based on evidence grids, is currently under development. For the experiments with elderly people, we are testing our first prototype of a non-invasive wireless multisensor device ($3.5 \times 3.5 \times 1.5$ cm), which measures room and skin temperature, GSR and acceleration.

Acknowledgements

This work is funded by Grant FP6-IST-045088 (SHARE-it). We would like to thank Prof. Kerstin Schill for her valuable comments on the early draft of this paper. The authors are also thankful to the anonymous reviewers for their useful suggestions and comments. Finally, we would like to acknowledge Christoph Budelmann's work on the wireless multisensor device.

References

Annicchiarico, R.; Cortes, U.; Federici, A.; Campana, F.; Barrue, C.; Martinez, A. B.; and Caltagirone, C. 2007. The Impact of Cognitive Navigation Assistance on People with Special Needs. In *IWANN 2007, San Sebastian, Spain*, 1060–1066. Springer-Verlag.

Anttonen, J., and Surakka, V. 2005. Emotions and Heart Rate while Sitting on a Chair. In *Proceeding of Computer Human Interaction 2005, CHI2005, Portland, Oregon, USA*.

Barrue, C.; Cortes, U.; and Annicchiarico, R. 2007. Shared Autonomy in Assistive Technologies. In *IWANN 2007, San Sebastian, Spain*, 1067–1073. Springer-Verlag.

Bradley, M. 2000. *Emotion and motivation*. New York: Cambridge University Press. 602–642.

Chatfield, C. 2003. *The Analysis of Time Series: An Introduction*. Texts in Statistical Science. Chapman & Hall/CRC, 6 edition.

Cichosz, J., and Slot, K. 2005. Low-dimensional feature space derivation for emotion recognition. In *INTERSPEECH-Eurospeech*, 477–480.

Desmet, P., and Hekkert, P. 2007. Framework of product experience. *International Journal of Design* 1(1):357–66.

Ekman, P. 1992. An argument for basic emotions. *Cognition and Emotion* 6:169–200.

Fragopanagos, N., and Taylor, J. 2005. Emotion recognition in human-computer interaction. *Neural Networks, Special Issue* 18:389–405.

Herbelin, B.; Benzaki, P.; Riquier, F.; Renault, O.; and Thalmann, D. 2004. Using physiological measures for emotional assessment: a computer-aided tool for cognitive and behavioural therapy. In *Proc. 5th Intl Conf. Disability, Virtual Reality and Assoc. Tech., Oxford, UK*.

Jang, J.-S. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics* 23:665–685.

Jones, C. M., and Jonsson, I.-M. 2005. Automatic recognition of affective cues in the speech of car drivers to allow appropriate responses. In *OZCHI '05: Proceedings of the 19th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction*, 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

Lang, P.; M.Greenwald; Bradley, M.; and Hamm, A. 1993. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30:261–273.

Mandel, C.; Huebner, K.; and Vierhuff, T. 2005. Towards an autonomous wheelchair: Cognitive aspects in service robotics. In *Proceedings of Towards Autonomous Robotic Systems (TAROS 2005)*, 165–172.

Mandryk, R. L., and Atkins, M. S. 2006. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *Elsevier. Int. J. Human-Computer Studies* doi:10.1016/j.ijhcs.2006.11.011.

McQuiggan, S.; Lee, S.; and Lester, J. 2006. Predicting user physiological response for interactive environments: an inductive approach. In *Proceedings 2nd Artificial Intelligence for Interactive Digital Entertainment Conference, AAAI Press*, 60–65.

Minguez, J., and Montano, L. 2004. Nearness diagram (nd) navigation: Collision avoidance in troublesome scenarios. *IEEE Transactions on Robotics and Automation* 20(1).

Panksepp, J. 1998. *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Posner, J.; Russel, J. A.; and Peterson, B. S. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* 17:715–734.

Russell, J., and Fehr, B. 1994. Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology* 67:186–205.

Russell, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39:345–356.

Simpson, R. C. 2005. Smart wheelchairs: A literature review. *IEEE Transactions on Robotics and Automation* 42(4):423–436.

Thrun, S.; Burgard, W.; and Fox, D. 2005. *Probabilistic Robotics*. The MIT Press.

Tomkins, S. S. 1962–1963. *Affect, imagery, consciousness Vols. I & II*. New York: Springer.

Vidrascu, L., and Devillers, L. 2007. Five emotion classes detection in real-world call center data: the use of various types of paralinguistic features. In *ParaLing'07, 16th International Conference on Phonetic Sciences, Saarbrcken, Germany*.

Witten, I. H., and Frank, E. 2005. *Data Mining. Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2 edition.