



Regression algorithm for emotion detection

Franck Berthelon, Peter Sander

► To cite this version:

Franck Berthelon, Peter Sander. Regression algorithm for emotion detection. CogInfoCom'13 - The 4th IEEE International Conference on Cognitive Infocommunicaitons, Dec 2013, Budapest, Hungary. hal-00908542

HAL Id: hal-00908542

<https://hal.science/hal-00908542>

Submitted on 24 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Regression algorithm for emotion detection

Franck Berthelon
Laboratoire I3S
Sophia-Antipolis, France
Email: berthelo@i3s.unice.fr

Peter Sander
Laboratoire I3S
Sophia-Antipolis, France
Email: sander@polytech.unice.fr

Abstract—We present two components of a computational system for emotion detection. PEMs (Personalized Emotion Maps) store links between bodily expressions and emotion values, and are individually calibrated to capture each person’s emotion profile. They are an implementation based on aspects of Scherer’s theoretical complex system model of emotion [1], [2]. We also present a regression algorithm that determines a person’s emotional feeling from sensor measurements of their bodily expressions, using their individual PEMs. The aim of this architecture is to dissociate sensor measurements of bodily expression from the emotion expression interpretation, thus allowing flexibility in the choice of sensors. We test the prototype system using video sequences of facial expressions and demonstrate the real-time capabilities of the system for detecting emotion. We note that, interestingly, the system displays the sort of hysteresis phenomenon in changing emotional state as suggested by Scherer’s psychological model.

I. INTRODUCTION

Recently there has been growing interest in computational emotion detection [3]. Among potential applications are games (both serious and otherwise), and human-machine interaction in general, promising systems which are more responsive to user behaviour. The work in this article falls squarely within the domain of cognitive info-communications - displaying and recognizing emotions can enable cognitive systems, artificial or natural, to better understand each other [4]. Emotion detection is something that seems natural for most people, but for a computational system the process is complicated by the huge variety of possible bodily expressions and interactions. Based on the definitions of Kleinginna [5] and Cowie [6], we propose a system able to detect episodic emotions in real time. The system’s interpretation algorithm is designed to be sensor agnostic, that is it will accept a wide variety of emotion expression measurements without relying on any specific sensor modalities.

A. Background

Computational emotion detection systems largely proceed via successive steps, from sensor measurement interpretation through an emotion model to the classification of emotional feeling as one from among a set of category choices (fear, joy, disgust, etc.) [7], [8], [9], [10]. In such classification systems the model is often a set of pre-existing templates or patterns to which the sensor output is matched. Our system has no such generic information; we employ a calibration step to personalize the system to individual users. This would make it more suitable for the applications mentioned above involving the willing cooperation of the participant, rather than for airport security, say, based on detecting nervousness from

facial expression. (Note however that lie-detectors also involve individual calibration.)

The obvious fact that people have no direct access to the emotions of others makes the emotion detection task doubly difficult for a computational system. First, we only have the bodily expression of emotion (ignoring language here) from which to infer the emotion that is being felt. Second, evaluating performance is complicated by the absence of a definitive objective ground truth against which to compare computed emotions. The second point influences the notion of calibrating or training the system with known examples, as we shall see below.

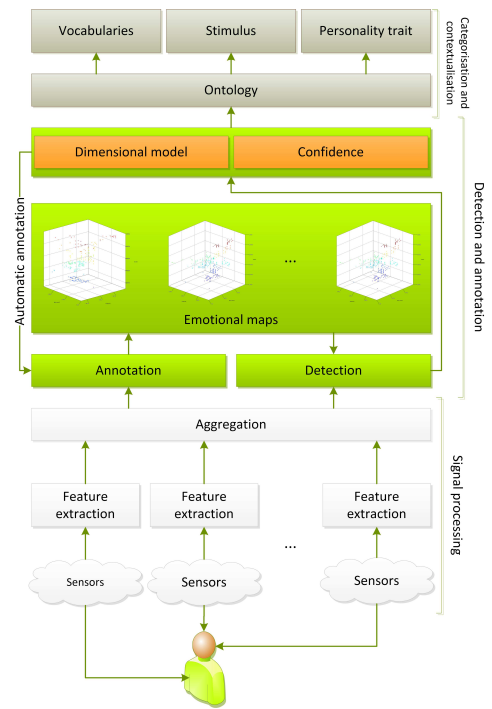


Fig. 1. Global architecture of our appraisal

An overview of the architecture of our system is presented in Figure 1. A key design consideration for our system was a separation of concerns between feature selection and feature handling of sensor data, which we consider as sensor-specific, from the step of information aggregation and interpretation which we would like to be sensor-independent. Interpretation of sensor data is realized by the combination of a PEM (Personalized Emotion Map), an implementation of Scherer’s complex-system model of emotion, and a regression algorithm allowing us to map emotion expression into a continuous

multi-dimensional emotion feeling space. This dimensional measure is weighted by the emotional context through an ontology to arrive at a categorial emotion label (this last step is described elsewhere[11]).

Different people react differently (with different bodily expression) to the same stimulus. As implied by the name, a PEM is personal to the individual. Its determination necessitates a calibration step using controlled stimuli before it can be useful for detecting that individual's emotional response to general stimuli.

Using Scherer's work, and more specifically his psychological model of emotion, as a theoretical base for our implementation allows us to treat emotion as a complex system consisting of multiple components. Handling individual components is relatively easy, but it's the global interactions between components which leads to the emergent behavior of the system. Further, this model has roots in work by Thom on the study of discontinuities, termed "catastrophes", to describe some emotion characteristics[12], particularly the transitions between emotions.

In Figure 2, hysteresis (a behaviour of Thom's catastrophes) is used to model the fact that for a same stimulus value we can't uniquely determine the emotion. We need to use previous emotion to know the direction of emotion evolution to determine which emotion is felt.

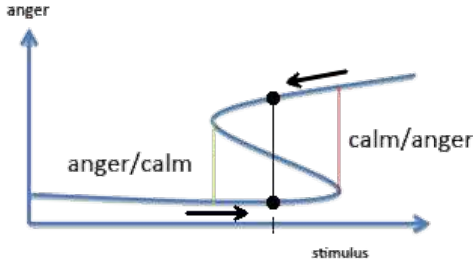


Fig. 2. Use of hysteresis in case of emotion modelization. The same stimulus value can give rise to different degrees of anger at the given point, depending on whether the point was approached by increasing or decreasing the stimulus (eg. using task where we can modulate difficulty to elicit boredom or anger). At the stimulus corresponding to the red vertical bar, under increasing stimulus, the system discontinuously jumps to the higher anger value (experiences a catastrophic increase in anger). Similarly at the green vertical bar, under decreasing stimulus, the system discontinuously drops to a lower value of anger.

In this paper, our aim is to present our modelization and especially the algorithm allowing us to interpret bodily expression into continuous emotion space. We choose a very simple two-dimensional emotion space, Russel's *arousal / valence* model [13]. This dimensional model is continuous, assigning numerical values along each dimension of the target space. In other work [11], we present mapping from this dimensional space into a categorial space of emotion labels and names.

We first present an overview of the PEM models, and focus on the algorithm used to exploit them. Second, we present an implementation of this work based on facial expression and an evaluation of this prototype computational system. The evaluation consists of a comparison between the system's determination of emotion from a video sequence and manual

annotation of the same sequence. Interestingly, the experiment suggests a hysteresis effect as predicted by Scherer's model.

II. PERSONALIZED EMOTIONAL MAP

A person feels emotion, but someone else can only see the bodily expression of that emotion. Our PEMs are designed as a link between emotion expression and emotion, calibrated for a particular person. They are used by our regression algorithm to predict emotion feeling, for that specific person. Bodily expression is measured by a variety of sensors, and a goal behind the development of the PEMs was that they should be suitable for use with different types of sensors. As well, they should be sufficiently lightweight to enable the algorithms to run in real-time.

As already mentioned, our model is inspired by the psychological work of Scherer, where complex system theory explains the emergence of emotions as a result of the combination of a number of bodily (and cognitive, we ignore those here, but see[11]) expressions. We describe below the formalism of the PEMs.

A. Model

Consider a heart-rate sensor - an increase in the sensor readings corresponds to an increase in heart-rate, which can indicate increased emotional arousal. However, feeling emotion is a complex phenomenon involving multiple body responses. According to emotion theory, each response by itself is emotionally non-specific [14], and in fact may arise independently of emotion. For example, increasing heart-rate can also be caused by exercise; pupil dilation can be caused by decreasing light. Emotion arises from the set of bodily responses rather than individual reactions. However, within the framework of real-time detection based on emotion expression sensors, it is not possible to implement this model directly. We thus modify Scherer's model, shown in Figure 3, to fit with our implementation constraints. Without getting into details here (we refer the interested reader to [11] for a more complete description), we cannot directly modify the control variables (power, facilitation) to influence the person's emotion, but we can modify the environment which will in turn indirectly vary the control variables. An example would be to increase obstructiveness by making gameplay increasingly difficult. Continuously modifying the control variables will trace a path over the behaviour surface. This will vary between individuals, and the PEMs are designed to capture individual emotion profiles.

We present here a brief formalization of the PEMs, and give implementation details below when describing the regression algorithm.

1) *Calibration*: PEMs are built during a calibration phase where:

- (i) the person is presented with varying stimuli;
- (ii) their reactions are indexed by $I_C \in [1, 2, 3, \dots]$ corresponding to a sequence indexing sensor measurements; and
- (iii) emotion values are associated with the sequence.

Somewhat more formally,

$$I_C \mapsto \mathbb{R}^M \text{ for } M \text{ sensors} \quad (ii)$$

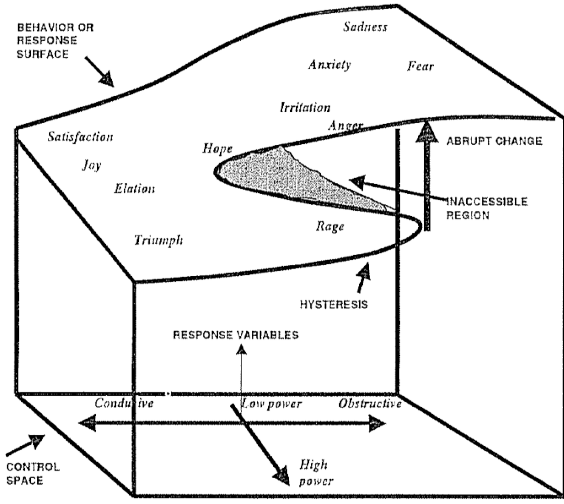


Fig. 3. Emotional surface from Scherer complex model with hysteresis.

This is not exactly what we want, rather we need a sequence of emotion values, eg, arousal / valence,

$$I_C \mapsto \mathbb{R}^N \text{ for } N \text{ emotion dimensions} \quad (\text{iii})$$

Now define the PEM as a composite mapping:

$$PEM : I_C \rightarrow \underbrace{\mathbb{R}^M \rightarrow \mathbb{R}^N}$$

Complete details of the parenthesized mapping which assigns emotion values to sensor measurements are not described here (see [11]), but in overview it register the expressions displayed by a person while feeling certain emotions.

Since I_C is fixed once the calibration is done, one can in fact consider a PEM as a set of points $PEM \subset \mathbb{R}^{M+N}$.

A calibrated PEM is presented in Figure 4, with only two sensor dimensions and one emotion dimension displayed for visibility. Dimensions M and N can be flexible, adding more sensors or emotion dimensions does not change the basic model formalism.

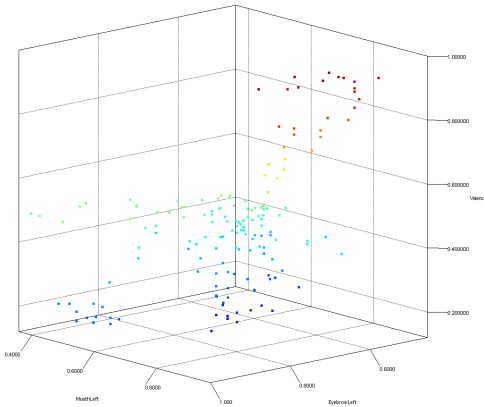


Fig. 4. Points represent valence values in PEM for sensor measurements. Temporal information not shown, and the data has been projected onto two sensor dimensions for visibility

For our approach to emotion detection, the calibration registers emotion as a temporal sequence, ie, a path in \mathbb{R}^{M+N}

space, in order to observe changes in emotion. A sequence then is a succession of states in the PEM corresponding to a given emotion. For example, to calibrate the system for joy, a new emotion sequence is created in the PEM corresponding to the expressions of the person while feeling joy, see Figure 4. PEMs may contain several sequences for the same emotion. Figure 5 shows some identified sequences in a PEM used during our experiments.

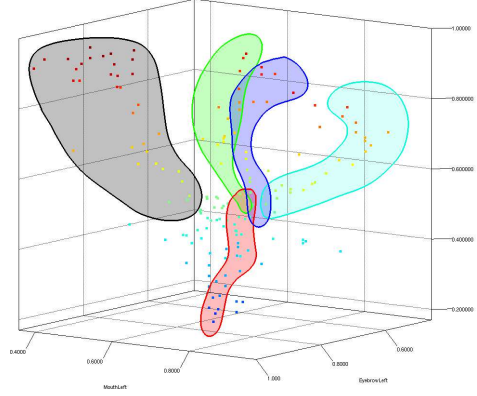


Fig. 5. Sequences in PEM for different emotions, eg, green - joy, grey - disgust, etc

2) *Detection*: Once the PEMs are calibrated, the task is to detect emotions when the person is confronted with different stimuli. Now we have another sequence of sensor measurements:

$$I_D \mapsto \mathbb{R}^M \quad (\text{ii})$$

and again we want a sequence of emotion values:

$$I_D \mapsto \mathbb{R}^N \quad (\text{iii})$$

where now the emotion values are to be determined automatically by the system. Detection is then defined as a mapping:

$$DETECTION : I_D \rightarrow \underbrace{\mathbb{R}^M \rightarrow \mathbb{R}^N}$$

The algorithm implementing the parenthesized mapping, which uses the calibrated PEMs, is described in the next section, and § III presents a demonstration of the system implementing this calibration / detection formalism.

B. Expression interpretation

The PEMs described above are used as the calibrated basis for our emotion detection algorithm. We use a KNN (k nearest neighbour) algorithm modified to work in a regression framework. We use regression to determine emotion in a continuous space \mathbb{R}^N , rather than a classification algorithm to pick a discrete emotion label or name based on previously-identified states.

Figure 6 shows the key point of our algorithm. A new sensor data point D_t , where $t \in I_D$ will become a point in $(M + N)$ -dimensional space. We already have M measured sensor values for D_t , we need to calculate the N unknown emotion values.

The algorithm determines the emotion values by using the $PEM \subset \mathbb{R}^{M+N}$. Conceptually (the actual implementation has a few tweaks to deal with initialization issues and such):

Regression algorithm for K = 6

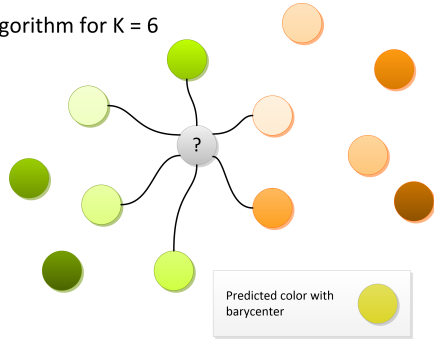


Fig. 6. Regression algorithm based on KNN

- (i) the N emotion values already found for the previous data point D_{t-1} are initially assigned as the values for D_t so that $D_t \in \mathbb{R}^{M+N}$;
- (ii) the KNN algorithm determines new emotion values for D_t as the weighted barycenter of the set of K nearest neighbour points in the PEM on each sequences.

The value of K , the number of neighbours to consider for each data point, is a parameter of the system that remains to be determined. We fix it empirically, as explained in § III-B1.

The emotion measure values are assigned a confidence value produced by the algorithm. Let K be the number of nearest neighbours, E_i the emotion state of neighbour i , S_j the unknown emotion state of the new data, and M the number of sensors. Then

$$confidence = \frac{1}{\sum_{i=1}^K \left(\sqrt{\sum_{j=1}^M (E_{ij} - S_j)^2} \right) / K} \quad (1)$$

As we said, our algorithm is executed on each PEM's sequences (see figure 4). We use confidence to determine the best result from each execution.

III. TESTING

A. Methodology

1) *Building the calibration database:* Using facial expression videos was chosen because the expressions are relatively easy to control for calibration, as opposed to heart-rate or GSR (Galvanic Skin Response, ie, sweat), and emotions are relatively easy for most people to determine from a sequence.

The experimental emotion expression test set was prepared from a video sequence of a subject (in fact an actor, in order to be able to generate canonical emotion expressions) displaying the six basic emotions identified by Ekman [15], see Figure 7. Between sequences displaying emotion, the subject's expression returned to neutral. Each "feature" of the face used in detection was considered as a separate sensor, eg, left corner of the mouth, middle of right eyebrow where we gather shifting as a descriptor for the facial expression and stored in our PEMs (see § II-A1).

To identify different emotion intensity, the subject displayed a range for each emotion from almost neutral to

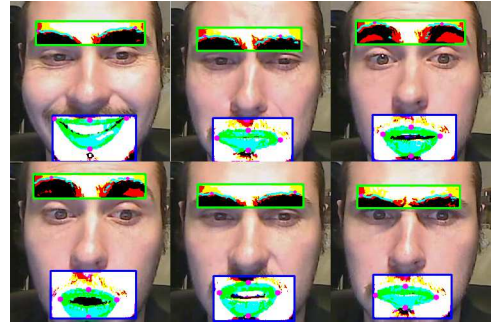


Fig. 7. Images from the video calibration sequences displaying Ekman's six basic emotions from top left to bottom right : joy, sadness, fear, surprise, disgust and anger.

extreme, see Figure 8. Each emotion sequence was repeated several times. All data were collected in one session to minimize potential problems with lighting and position.

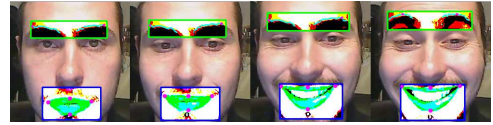


Fig. 8. Images from the video calibration sequences displaying emotion intensity variation for joy

2) *System performance:* To quantify the emotion detection performance of the system, we define a notion of *deviation*. This corresponds to the difference between emotion detection by hand in the video sequences and automatic detection by the system. We calculate average deviation by component (arousal and valence), by emotion and for the whole set of emotions.

We define the deviation for each emotion component as follows:

$$\Delta = \frac{\sum_i^n (anno_{man_i} - anno_{auto_i})}{n} \quad (2)$$

Where *anno* are the emotion values defined by human (*anno_{man}*) or automatically by the system (*anno_{auto}*).

For the average deviation of an emotion we use:

$$\mu_{emotion} = \frac{\sum_i^n |\Delta_i|}{n} \quad (3)$$

The average deviation is defined by:

$$\mu = \frac{\sum_i^n |\mu_{emotion_i}|}{n} \quad (4)$$

The first measure is used to observe system variation as a function of the emotion to detect. The second gives some idea of the robustness of the system to certain internal parameter changes, ie, the neighbourhood size parameter K , or the number of states at system initialization.

Figure 9 represents the methodology used for quantifying system performance. Only the first bloc needs the subject present, the rest is *a posteriori* processing. This allowed us to determine performance while performing internal system modifications.

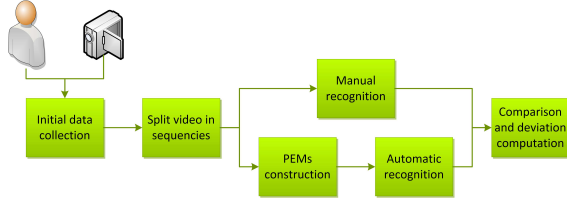


Fig. 9. Experimentation methodologie

B. Results

1) *Determining the value of K* : The graph in Figure 10 shows the average deviation for each emotion and the overall average deviation for different values of K .

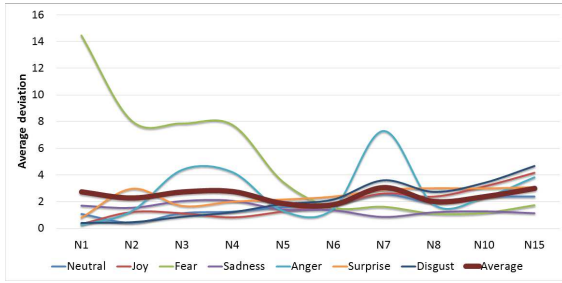


Fig. 10. Deviation function of K parameter

The figure shows the drop in deviation for fear (green curve) with increasing K , up to about $K = 6$. Thereafter increasing K leads to generally increasing deviation. We have therefore fixed the value $K = 6$ for experiments as this seemed a reasonable compromise.

2) *Emotion detection in the test video*: After calibration, we tested the system on the whole facial expression video, containing 5757 images representing sequences of the six basic emotions. In all, each emotion was repeated a variable number of times as there were 17 separate sequences. Figure 11 presents the results obtained; blue is valence and red is arousal. The figure shows that there were indeed roughly 17 sharp changes in the labeling corresponding to the 17 different sequences.

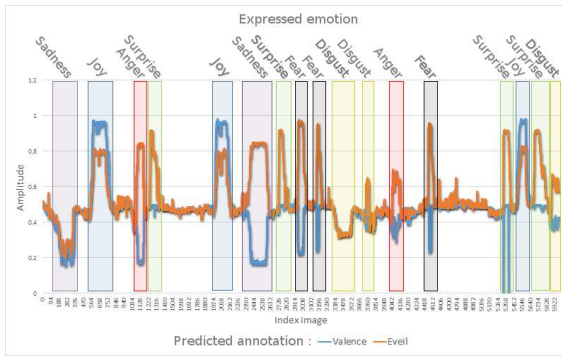


Fig. 11. Automatic emotion annotation based on facial expression

Looking closely at the results we have identified two emotion mis-identification errors, between anger and disgust. In all other cases the automatic labeling is coherent with the emotion expressed by the subject.

3) *Hysteresis observation during emotion detection*: Basing our emotion detection system on the complex system model proposed by Scherer[1], [2] is not without risk. As far as we know, this is the first attempt at turning the psychological model into a computational implementation. One of our goals was to experimentally test the underlying theoretical hypotheses of the model. Here we are not proposing a full validation of the model, but our system does display behaviour which suggests that it is performing as predicted by the model. This is the notion of emotion hysteresis. The theory predicts that under variable anger-producing stimulus, the point at which a person would jump from calm to anger with increasing stimulus is not the same point at which they would return from anger to calm for decreasing stimulus. This insight is easy to show diagrammatically (see Figure 2), but the multi-dimensionality of our PEMs renders this difficult. As well, the effect is easily shown for a response curve or surface containing a fold, but the PEMs consist of a cloud of points.

However, we have extracted an automatically labeled sequence from the test video where the emotion displayed went from calm to joy and back to calm again. Intermediate states were labeled during the sequence. This is shown graphically as the path in Figure 12 starting with calm at the blue dot and returning to calm at the black dot. The striking observation is that the return to calm follows a somewhat different series of intermediate states than does the path toward joy.

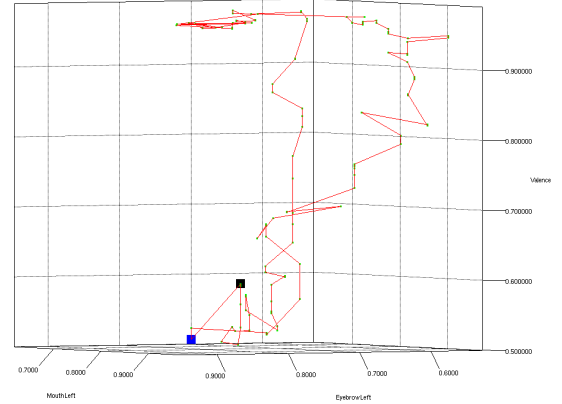


Fig. 12. Automatic annotation based on facial expression. The emotion paths from neutral to joy and joy to neutral are different

To show that this is not just an artefact of the system, we duplicated and reversed the neutral-to-joy subsequence to produce a symmetric sequence. Running it through our system produced the results shown in Figure 13, where the to and fro paths are much closer. We are planning work to further investigate this phenomenon.

C. Discussion

The experiment tested the performance of our system in interpreting emotion expression in a video sequence of facial images by comparing its emotion assessment with that of a human. Calibrating the system allowed us to tune certain parameters for more robust results, eg, fixing the value of $K = 6$ in the KNN algorithm. We are quite satisfied with the KNN

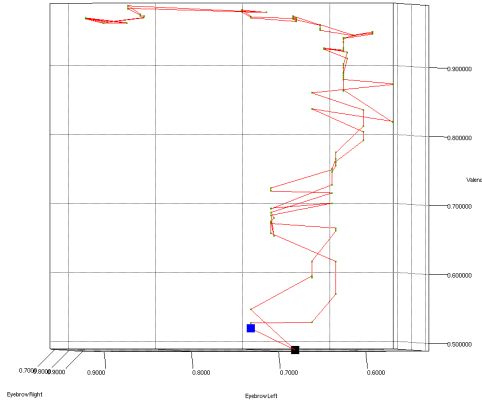


Fig. 13. Automatic annotation based on facial expression, with a symmetric sequence. The emotion paths from neutral to joy and joy to neutral are similar

based regression algorithm as it seems well-adapted to our case. While perhaps not highly considered for classification, it enabled our algorithm to run at between 8 and 12 images per second on a run-of-the-mill laptop. This is sufficient for our ends as emotions tend to last on the order of seconds. At this speed, the system can follow the evolution and changes in emotions through the PEMs.

When run over a video containing 17 sequences of distinct emotion expressions, with transition through a neutral emotion expression, the system only missed identifying two sequences. The system has difficulty deciding between anger and disgust. The images expressing these emotions are actually quite similar, but perhaps with more and better calibration the system would be able to distinguish them. We suspect that adding further sensors like heart-rate would also lead to better performance. This probably applies to other emotions as well and is the subject of future work.

The experiments also suggest evidence for emotion hysteresis, an asymmetry in the transition between emotions. This effect is hypothesized by Scherer's complex system model of emotions which forms the basis model for our implementation. The effect is not present in an artificially created symmetric sequence, and so seems not to be an artefact of our system implementation.

IV. CONCLUSION

We have presented an architecture for a system for determining emotion from multi-sensor input. A design goal was to abstract specific sensor information away from the algorithmic heart of the system. The only constraint on the sensors is that they provide continuous measures. We introduced Personalized Emotion Maps (PEMs) to map sensor input into emotion values. The model behind our maps is from a psychological theory of emotion based on complex system theory. Once calibrated for individual responses (hence the name), PEMs are used by a KNN based regression algorithm to associate emotion values with sensor input.

An experiment on emotion determination from a video sequence of facial images demonstrated the performance of the

system. The system performed remarkably well, only missing out on two of 17 different emotion sequences in the video. The system also demonstrated a hysteresis effect in emotion transition, as predicted by the model.

Currently the system is just a prototype and quite rough around the edges. Much work remains to be done to refine the different system components. More work is also needed to determine the variability in the PEM calibration over a larger range of stimuli than presented in this paper. We are exploring using personal emotion ontologies to take into account some of the expected variability in emotion measurements between users faced with the same stimulus[11]. We are also studying the emotions solicited within a same user by different types of stimuli which are expected to provoke similar emotions, eg, satisfaction or joy at either succeeding at a computer game, or while watching a favourite movie sequence.

ACKNOWLEDGMENT

The authors wish to thank the Stonetrip company, and the PACA region for their support of this work.

REFERENCES

- [1] K. R. Scherer, "Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes," 2000.
- [2] —, "Emotions are emergent processes: they require a dynamic computational architecture," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3459–3474, 2009.
- [3] R. W. Picard, "Affective computing," *Pattern Recognition*, 1995.
- [4] P. Baranyi and A. Csapo, "Cognitive infocommunications: Coginfo-com," in *11th IEEE International Symposium on Computational Intelligence and Informatics*, 2010.
- [5] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions with suggestions for a consensual definition," *Motivation and Emotion*, 1981.
- [6] R. Cowie and M. Schröder, *Piecing together the emotion jigsaw*, ser. Lecture Notes in Computer Science. Springer Verlag, 2005, vol. 3361, pp. 305–317, book Title: Machine Learning for Multimodal Interaction, First International Workshop, MLMI 2004, Martigny, Switzerland, June 21–23, 2004, Revised Selected Papers.
- [7] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [8] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," *Neural Computing & Applications*, vol. 9, no. 4, pp. 290–296, 2000.
- [9] D. O. Bos, "Eeg-based emotion recognition," 2006.
- [10] M. Li and B.-L. Lu, "Emotion classification based on gamma-band eeg," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 1223–1226.
- [11] F. Berthelon, "Modélisation et détection des émotions à partir de données expressives et contextuelles," Ph.D. dissertation, University of Nice-Sophia Antipolis, In progress.
- [12] R. Thom, *Stabilité structurelle et morphogénèse*. WA Benjamin Reading, Mass, 1972.
- [13] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, 1980.
- [14] W. B. Cannon, "The james-lange theory of emotions: A critical examination and alternative theory," *American Journal of Psychology*, vol. 39, pp. 106–124, 1927.
- [15] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in *Nebraska symposium on motivation*. University of Nebraska Press, 1971.