The Smart Sensor Integration Framework and its Application in EU Projects

Johannes Wagner, Frank Jung, Jonghwa Kim, Thurid Vogt and Elisabeth André

Multimedia Concepts and Applications, University of Augsburg
Universitätsstr. 6a, D-86159 Augsburg, Germany

Abstract. Affect sensing by machines is an essential part of next-generation human-computer interaction (HCI). However, despite the large effort carried out in this field during the last decades, only few applications exist, which are able to react to a user's emotion in real-time. This is certainly due to the fact that emotion recognition is a challenging part in itself. Another reason is that so far most effort has been put towards offline analysis and only few applications exist, which can react to a user's emotion in real-time. In response to this deficit we have developed a framework called Smart Sensor Integration (SSI), which considerably jump-starts the development of multimodal online emotion recognition (OER) systems. In this paper, we introduce the SSI framework and describe how it is successfully applied in different projects under grant of the European Union, namely the CALLAS and METABO project, and the IRIS network.

1 Introduction

Next generation human-computer interaction (HCI) claims to analyze and understand the way users interact with a system in a more sophisticated and smarter way than traditional systems do. No longer should it be the user who adapts to the system, but the system that adjusts itself to the user. This requires the system to be not only aware of the users' goals and intensions, but also their feelings and emotions. For this reason, during the last decade, plenty of methods have been developed to detect a user's emotions from various input modalities, including facial expressions [12], gestures [2], speech [9], and physiological measurements [7]. Also, multimodal approaches to improve recognition accuracy are reported, mostly by exploiting audiovisual combinations [1].

To date, however, most of the systems have been developed for offline processing and are not yet ready to be used under real-time conditions. This, of course, hampers their usefulness in practical applications. We believe that this is due to the varied difficulties real-time capability implies. On the one hand, an online system needs to deal with additional requirements, such as automatic segmentation, normalization issues, or the constraint to build on low-cost algorithms. On the other hand, there are certain implementation hurdles arising from the parallel processing of the input modalities, *e.g.* sensor data must be permanently captured and processed, while at the same time classification has to be invoked on detected segments. While for the offline analysis of

Wagner J., Jung F., Kim J., Vogt T. and André E. (2010).
The Smart Sensor Integration Framework and its Application in EU Projects.
In *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications*, pages 13-21
DOI: 10.5220/0002813900130021

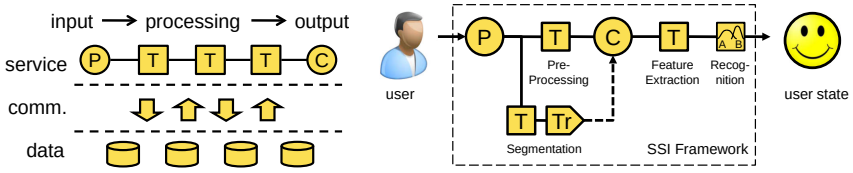

Fig. 1. The left part of the figure shows the three-layered architecture of the SSI framework. On the right side a basic configuration of an emotion recognition system implemented with SSI is displayed. In the charts P stands for provider, T for transformer, Tr for Trigger and C for consumer.

emotional corpora a number of powerful tools are available, such as Anvil[1] for data annotation, Matlab for signal processing and Weka[2] for classification, only little support is given for Online Emotion Recognition (OER) systems.

This paper reports on a framework called Smart Sensor Integration (SSI), which we have developed at our lab in order to support the building of OER systems. Originally designed to tune our own efforts on offline processing to the more challenging task of online processing, we have now made it available to the public, hoping to contribute to the implementation of OER systems in the future. In the following, we will shortly introduce the concept and architecture of the SSI framework, followed by a description how SSI is successfully applied in a number of projects funded by the European Union.

2 Smart Sensor Integration (SSI)

Online Emotion Recognition (OER) deals with two tasks: a learning phase, which involves data acquisition and training of a model from the data, and setting up an online recognizer, which is able to track a user's affective state. A framework meant to support the creation of OER systems must consider both tasks. To this end, SSI, which in the first place is a framework made of tools for setting up an online recognition pipeline, also includes a graphical user interface for data acquisition and training.

2.1 Building Pipelines for Online Emotion Recognition

As depicted in the left part of Figure 1 the SSI framework has a three-layered architecture. The two lower levels, called data and communication layer, are responsible for data buffering and access. They temporarily store the processed signals. A synchronization mechanism handles the simultaneous access to the data and enables us to request snapshots of different signals for the same time slot, *e.g.* fetch video frames and the

[1] Anvil is a video annotation tool offering hierarchical multi-layered annotation driven by user-defined annotation schemes.
http://www.anvil-software.de/

[2] Weka is an open source software, which offers a large collection of machine learning algorithms for data mining tasks.
http://www.cs.waikato.ac.nz/ml/weka/

according portion of audio. An internal clocking mechanism takes care of data synchronization and restores it if necessary.

A developer, however, only has to deal with the top layer of the framework, called service layer. It allows the easy integration of sensors, processing algorithms, triggers, and output components to a signal processing pipeline capable of live input. The advantage for the developer is that he is not struggled with the usual problems online signal processing involves. To build the pipeline he can either integrate own code, or choose from a large number of available components.

Available components fall into three categories: Components related to data segmentation responsible to automatically detect chunks of activity. Components, which apply necessary pre-processing and extract from the detected chunks meaningful features that express the changes in the affective state of the user. And finally, the models, which are used to map from a continuous feature space to discrete emotion categories.

The right part of Figure 1 shows the basic configuration of an emotion recognition system implemented with SSI. A further advantage of using a generic framework like SSI is the great amount of flexibility and reusability that is gained. Since the processing units work independently of each other, they can be easily re-assembled in order to experiment with different kind of settings or to fit new requirements. In an earlier paper we show examples, how a basic OER system can be stepwise extended to a more complex one, *e.g*. by fusing information from multiple modalities [11]. The SSI framework has been published under LGPL license and can be freely obtained from https://mm-werkstatt.informatik.uni-augsburg.de/ssi.html.

2.2 Data Acquisation and Classifier Training

The model that a classifier uses to map continuous input to discrete categories is initially unknown and has to be learned from training data. This task falls into two main parts. The first part, referred to as data acquisition, involves the collection of representative training samples. Here, representative means that the picked samples should render the situation of the final system as accurate as possible. The second part concerns the actual training of the classifier. Basically, this is related to the problem of training a model that gives a good separation of the training samples, but at the same time is generic enough to achieve good results on unseen data.

Offline classification is usually evaluated on a fixed training and test set collected under similar experimental setting and by tuning the model parameters until they yield optimal recognition results. In contrast, the success of an online system depends on the ability to generalize on future data, which is not available to evaluate the classifier. Hence, special attention has to be paid that the training data is obtained in a situation, which is similar to the one it will be used in. To this end, we have developed a graphical user interface (GUI) on top of SSI, which gives non-experts the possibility to record emotional corpora and train personalized classifiers, which can be expected to give considerably higher accuracy than a general recognition system.

The GUI will be introduced by means of a concrete application within the CALLAS project in Section 3.1.

3 SSI in Practical Applications

A main motivation for building the SSI framework has been our participation in different EU projects, which are concerned with the analysis and development of novel user interaction methods. They all require to some extent online detection of the user's affective state. In the following we explain how SSI is applied to this task.

3.1 The CALLAS Project

The CALLAS[3] (Conveying Affectiveness in Leading-edge Living Adaptive Systems) project aims to develop interactive art installations that respond to the multimodal emotional input of performers and spectators in real-time. Since the beginning of the project a various number of showcases have already been created, such as the E-Tree[4], which is an Augmented Reality art installation of a virtual tree that grows, shrinks, changes colours, etc. by interpreting affective multimodal input from video, keywords and emotional voice tone, or Galassie[6] by Studio Azzurro, which creates stylized shapes similar to galaxies depending on the users' emotional state detected from the voice.

One of the research questions raised by CALLAS is the interpretation, understanding, and fusion of the multimodal sensor inputs. However, since multimodal data corpora with emotional content are rather rare, effort was made to create a setup, which simplifies the task of data acquisition. In one experiment, which targets the mapping between gestures/body movements and emotion, and their relation to other modalities, such as affective speech and mimics, user interaction is captured with two cameras, one focused on his head and one on his whole body, and a microphone near the head. Additionally the users interact with different devices, such as Nintendo's Wii Remote or a data glove by HumanWare[4]. To elicit the desired target emotion a procedure inspired by the Velten emotion induction method is used: first, a sentence with a clear emotional message is displayed and the user is given sufficient time to read it silently. Then the projection turns blank and the user is asked to express the according emotion through gesture and speech. It is up to the user to use own words or to say something, which is similar to the displayed sentence. Figure 2 illustrates the setting.

A main challenge regarding the experimental design is the proper synchronization between the different modalities. This, however, is an important requirement for the further analysis of the data. To obtain synchronized recordings we use the SSI framework. SSI already supports common sensor devices, such as webcam/camcorder, microphone and the Wii Remote, and can also record from multiple devices of the same kind. To connect more exotic devices, such as the data glove, SSI offers a socket interface, which can be used to grab any signal stream and feed it into the processing pipeline. SSI takes care of the synchronization by constantly comparing the incoming signals with an internal clock. If a sensor breaks down or does not deliver the appropriate amount of data, SSI compensates the lack in order to keep the stream aligned with the other channels. If it is not possible to capture all sensors with the same machine - which was actually the case with the two high-quality cameras, due to the vast amount of data they produce

[3] http://www.callas-newmedia.eu
[4] http://www.hmw.it/

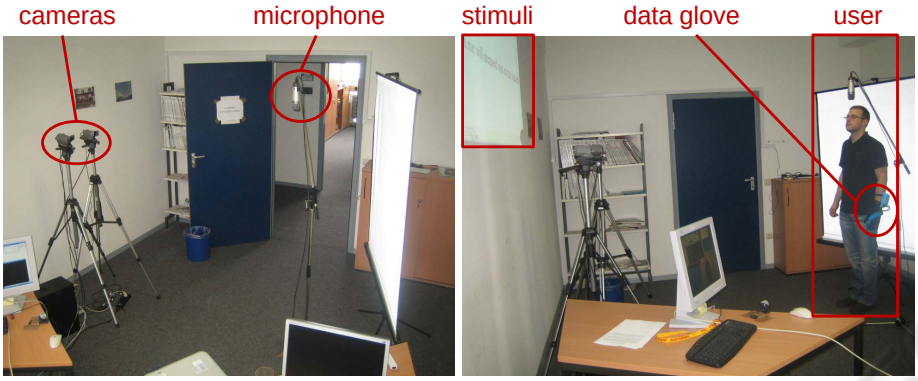

Fig. 2. The Figure shows the setting of an affective gesture recognition experiment carried out in the CALLAS project. Users are recorded from two cameras, one microphone, and additional interaction devices such a data glove.

- SSI offers the possibility to synchronize recordings among several computers using a broadcast signal, which is sent through the network.

For the accomplishment of the experiment we use the graphical interface of SSI, called SSI/ModelUI (see Figure 3). It works on the top of SSI and allows the experimenter to display a sequence of HTML documents, which contain the according instructions or stimuli. When the recording is finished it is added to the database. However, the functionality of the tool goes beyond this. During the recording SSI already tries to detect the interesting parts in the signals, e.g. when a user is talking or performing a gesture. These events are stored and can be reviewed together with the videos and the other raw signals. An annotator can now crawl through the events and adjust these pre-annotations. Finally, the tool automatically extracts feature vectors for the labelled segments and uses them to train a model for online classification. This way, the tool combines the tasks of recording, annotation and training in one application.

At the moment, we have conducted the presented experiment in two countries: Greece and Germany. We have recorded 30 subjects (10 Greek, 20 German) of which half were male and the other half female. The total length of recorded data sums up to almost 9h. The study will be repeated in Italy and possibly in countries of other partners. The analysis of the corpus has recently started and first results can be soon expected. The corpus will also allow us to analyze similarities/differences between individuals and between cultural groups in terms of selected features, recognition results and use of modalities.

3.2 The METABO Project

METABO[5] is a European collaborative project with the aim to set up a platform for monitoring the metabolic status in patients with, or at risk of, diabetes and associated

[5] http://www.metabo-eu.org/

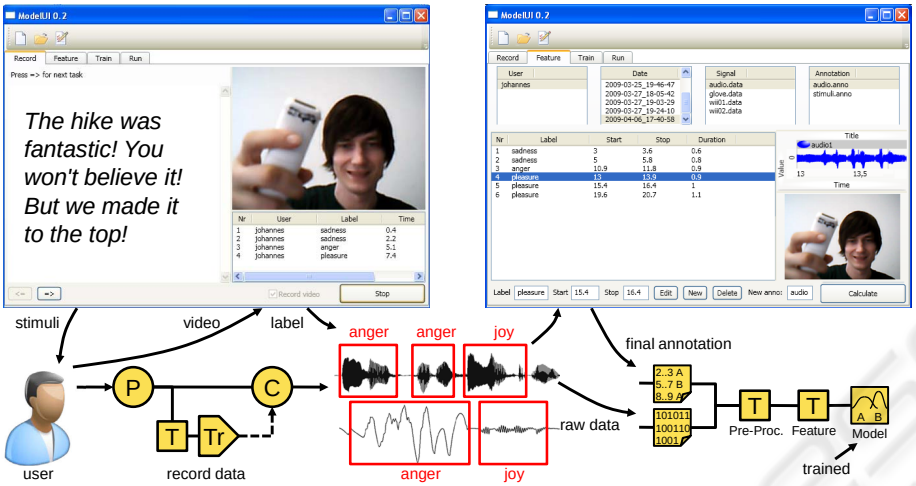

Fig. 3. The Figure shows two screenshots of the graphical interface we have developed on top of SSI. It helps the experimenter to record the user interaction (left image) and let him review and annotate the recordings (right image). It also leads through the other steps of the training procedure. The GUI uses a DLL to communicate with SSI (see sketches below the screenshots).

metabolic disorders. The platform serves as bridge between physicians and patients to exchange information, but at the same time also provides recommendations for their clinical treatment. Recommendations are generated individually based on the patient's metabolic behaviour model, which is learned from the patient's history, and the current context derived from the patient's physiological state. An essential requirement for such a system is the online acquisition and the prompt analysis of user data measured from different sensor devices. For this purpose a special case study is carried out, called the in-vehicle hypoglycemia alerting system (IHAS).

IHAS is an emotion monitoring system, which analyses a driver's behaviour and emotional state during driving. Based on the measurements the system predicts hypoglycemia events and alerts the driver. The setting is motivated by the critical role fluctuant emotions play for diabetic drivers [8]. To derive the emotional state, the driver is equipped with several biosensors, including electromyogram (EMG), galvanic skin response (GSR), electrocardiogram (ECG), and respiration (RSP). The captured data is analyzed using an online recognition component implemented with the SSI framework.

In our previous studies we have mainly focused on the offline analysis of physiological data, where we have tested a wide range of physiological features from various analysis domains including time, frequency, entropy, geometric analysis, sub-band spectra, multi-scale entropy, and HRV/BRV. When we applied these features to different data sets recorded at our lab, we were able to discriminate basic emotions, such as joy, anger or fear, with an accuracy of more than 90% [10, 7]. For the sake of real-time analysis, large parts of the code that has been originally developed in Matlab, were ported to C++ and incorporated into the SSI framework and offline algorithms were replaced by corresponding real-time versions. That recognition results do not necessarily drop when

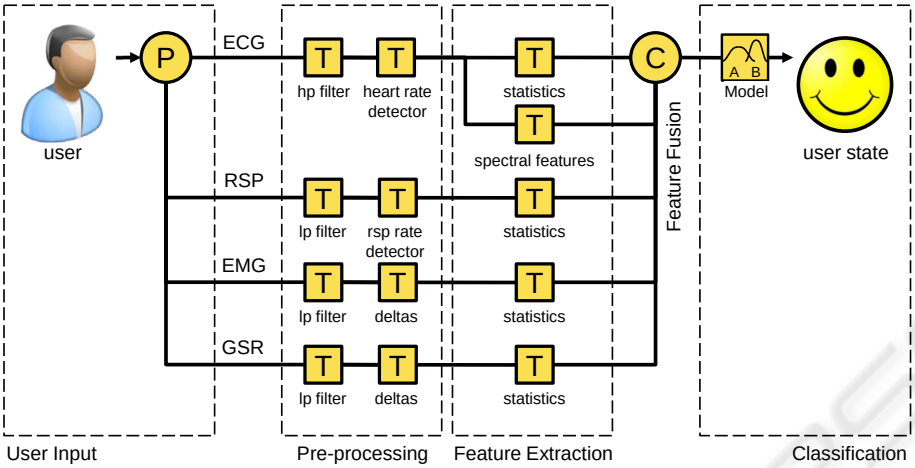

Fig. 4. State flow of the online emotion recognition component developed with SSI as part of the METABO project: each sensor channel is first processed individually before features are finally combined to a single feature vector. The combined feature vector is analyzed by the recognition module.

a generic set of recursively calculated real-time features is used instead of specialized offline features has been shown by Hönig *et al*. [5]. Many of the new real-time features are oriented to their proposed feature set. Figure 4 shows the state flow of the online recognition component we have developed with the SSI framework and which has been integrated into the METABO platform.

3.3 The IRIS Project

The IRIS[6] (Integrating Research in Interactive Storytelling) project is concerned with the development of novel technologies for interactive storytelling. Recognition of affect is one of the novel techniques to be integrated into virtual storytelling environments. One of the showcases developed by Teesside University, EmoEmma[3], is based on Gustave Flaubert's novel "Madame Bovary". Here, the user can influence the outcome of the story by acting as one of the characters and their interaction mode is restricted to the emotional tone of their voice.

In a first step, we have integrated the EmoVoice system [9] into the SSI framework. EmoVoice has been developed at our lab as a tool for the recognition of affective speech. It provides tools for acoustic feature extraction (no semantic information is used) and building an emotion classifier for recognizing emotions in real-time. In total, a set of 1451 features can be derived from pitch, energy, voice quality, pauses, spectral and cepstral information as conveyed in the speech signal. The performance of the system has been evaluated on an acted database that is commonly used in offline research (7

[6] http://iris.scm.tees.ac.uk

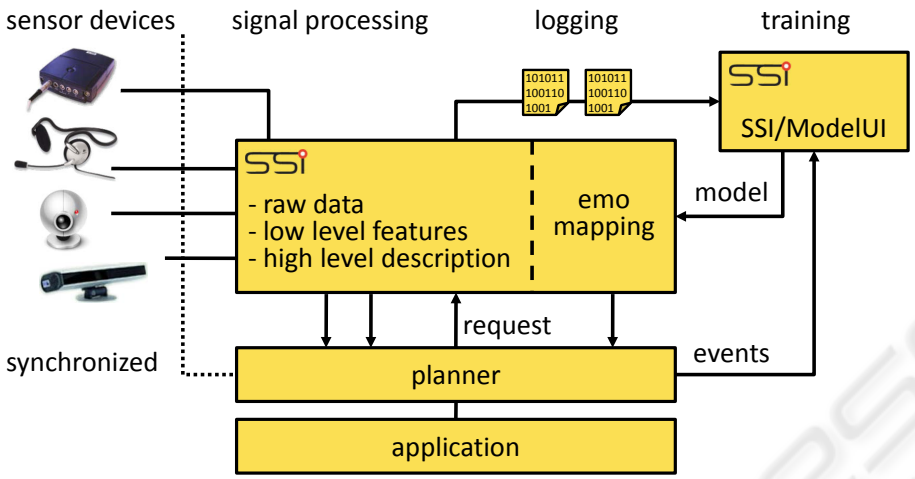

Fig. 5. In the IRIS project, which is about next-generation interactive storytelling, user interaction is measured with different sensor devices that are synchronized and pre-processed through the SSI framework. The recognized user behaviour is finally used by the planner to decide which path is taken in the progress of the story.

emotion classes, 10 professional actors) and on a speech database we recorded for online classification (4 emotion classes, 10 German students) and achieved an average recognition accuracy of 80% and 41% respectively.

With the integration we have set the ground for a more complex analysis of the user interaction as now additional sensors can be added with minimal effort. To analyze the user's affective and attentive behaviors, we use SSI in two ways: firstly, it offers the possibility to collect synchronized sensor data from users interacting with the system. In order to get realistic data the system response will be simulated at this point. Afterwards we analyze the recordings and based on the observations we use SSI to implement a pipeline that extracts the observed user behaviour in real-time. This information is then used by the planner to automate the system response. The architecture of the system is shown in Figure 5.

4 Conclusions

In the past pages, we have introduced our approach to contribute the building of OER systems, a framework called Smart Sensor Integration (SSI). First, we have discussed the multiple problems concerned with the implementation of such systems and how SSI faces them. After a short introduction of the framework architecture, we have moved on to explain how SSI is applied in a number of projects funded by the EU. SSI is freely available under LGPL license from the following address:
https://mm-werkstatt.informatik.uni-augsburg.de/ssi.html.

Acknowledgements. The work described in this paper is funded by the EU under research grants CALLAS (IST-34800), IRIS (Reference: 231824) and Metabo (Reference: 216270).

References

1. Bailenson, J.; Pontikakis, E.; Mauss, I.; Gross, J.; Jabon, M.; Hutcherson, C.; Nass, C.; John, C.: Real-time classification of evoked emotions using facial feature tracking and physiological responses. *Int'l Journal of Human-Computer Studies*, 66(5):303-317, 2008.
2. Caridakis, G.; Raouzaiou, A.; Karpouzis, K.; Kollias, S.: Synthesizing gesture expressivity based on real sequences. In *Proc. of LREC Workshop on multimodal corpora: from multimodal behaviour theories to usable models*, 2006.
3. Charles, F.; Pizzi, D.; Cavazza, M.; Vogt, T.; André, E.: Emotional input for character-based interactive storytelling. In *The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Budapest, Hungary, 2009.
4. Gilroy, S. W.; Cavazza, M.; Chaignon, R.; Mäkelä, S.-M.; Niranen, M.; André, E.; Vogt, T.; Urbain, J.; Billinghurst, M.; Seichter, H.; Benayoun, M.: E-tree: emotionally driven augmented reality art. In *Proc. ACM Multimedia*, pages 945-948, Vancouver, BC, Canada, 2008. ACM.
5. Hönig, F.; Wagner, J.; Batliner, A.; Nöth, E.: Classification of user states with physiological signals: On-line generic features vs. specialized feature sets. In *Proc. of the 17th European Signal Processing Conference (EUSIPCO-2009)*, 2009.
6. Jacucci, G. G.; Spagnolli, A.; Chalambalakis, A.; Morrison, A.; Liikkanen, L.; Roveda, S.; Bertoncini. M.: Bodily explorations in space: Social experience of a multimodal art installation. In *Proc. of the twelfth IFIP conference on Human-Computer Interaction: Interact*, 2009.
7. Kim, J.; André, E.: Emotion recognition based on physiological changes in music listening. *IEEE Trans. Pattern Anal. and Machine Intell.*, 30(12):2067-2083, 2008.
8. Kim, J.; Ragnoni, A.; Biancat, J.: In-Vehicle Monitoring Of Affective Symptoms for Diabetic Drivers. In *Proc. of the Int. Conf. on Health Informatics (HEALTHINF)*, 2010. [in press]
9. Vogt, T.; André, E.; Bee, N.: A framework for online recognition of emotions from voice. In *Proc. of Workshop on Perception and Interactive Technologies for Speech-Based Systems*, Kloster Irsee, Germany, 2008.
10. Wagner, J.; Kim, J.; André, E.: From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In *Proc. IEEE ICME 2005*, pp. 940-943, Amsterdam, 2005.
11. Wagner, J.; André; E., Jung, F.: Smart sensor integration: A framework for multimodal emotion recognition in real-time. *In Affective Computing and Intelligent Interaction (ACII 2009)*, 2009.
12. Zeng, Z.; Pantic, M.; Roisman, G I.; Huang, T S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(1):39-58, 2009.