

An environment for studying visual emotion perception

Davide Carneiro^{1,2}, Helder Rocha¹, and Paulo Novais¹

¹ CIICESI, ESTG, Polytechnic Institute of Porto
Felgueiras, Portugal

² Algoritmi Center/Department of Informatics, University of Minho
Braga, Portugal
{dcarneiro,pjon}@di.uminho.pt, pg27737@alunos.uminho.pt

Abstract. Visual emotion perception is the ability of recognizing and identifying emotions through the visual interpretation of a situation or environment. In this paper we propose an innovative environment for supporting this type of studies, aimed at replacing current pencil-and-paper approaches. Besides automatizing the whole process, this environment provides new features that can enrich the study of emotion perception. These new features are especially interesting for the field of Human-Compute Interaction and Affective computing as they quantify the effects of experiencing different emotional dimensions on the individual's interaction with the computer.

Keywords: Visual Emotion Perception, Behavioural Biometrics, Keyboard Dynamics

1 Introduction

Emotions are one of the most interesting Human mechanisms and serve very important and diverse functions, namely by prompting one for action, providing information about a situation or allowing a more efficient communication with other individuals [1].

While the study of emotion *per se* and of its role in inter-human relationships is not new, in the last years a new field of study emerged that is devoted to the study of emotions in the context of Human-Computer Interaction: Affective Computing[2]. Specifically, this field aims at the development of systems that can recognize and respond to Human emotions.

In this paper we detail the development of a system to facilitate the study of visual emotion perception. Moreover, the system also aims to measure the influence of these emotional dimensions on Human-Computer Interaction.

The goal is, on the one hand, to facilitate data collection, allowing simultaneous data collection processes and providing the researcher community with access to relevant data in a structured manner. On the other hand, we aim to provide new variables that can enrich this type of studies, eventually leading to new and relevant findings. Namely, this system combines behavioural biometrics

and accelerometers placed on the keyboard, to characterize, in a rich way, the interaction of the participant with the computer.

This is, indeed, the key innovative aspect of this system: the non-intrusive collection of new types of information from the environment, that can more thoroughly characterize the participant's state and significantly enrich the work of the researcher and widen the potential findings.

Despite the specific scope of the current work, the developed environment can undoubtedly be used with other aims (e.g. stress assessment, fatigue detection).

1.1 System Operation

The proposed system allows researchers to select (using the Administration dashboard) from a large set of emotionally-evocative color photographs from the International Affective Picture System (IAPS)[3]. The IAPS is a collection of normatively rated affective stimuli, considering three dimensions: affective valence (ranging from pleasant to unpleasant), arousal (ranging from calm to excited) and dominance (ranging from controlled to in-control).

The researcher selects the desired stimuli and creates and configures a new study, consisting of a sequence of photographs. These photographs are sequentially shown to the participant, either in a pre-determined order or randomly.

The participant is asked to describe (typing in the keyboard) how she/he actually feels while watching each of the photographs. The participant is allowed to look at the photograph while typing and there is no time-limit for this task. In the following screen the participant is also asked to rate each picture in terms of valence, arousal and dominance. A 9-point Likert-scale is used for this rating with the following labels: valence (1 - unpleasant, 9 - pleasant), arousal (1 - calm, 9 - excited) and dominance (1 - dominated, 9 - in control).

After the participant provides this feedback, there is a 30-seconds interval in which no photograph is shown (only a white rectangle is visible in the space of the photograph) before advancing to the next one. This is done to "reset" the emotional state of the participant between each two photographs.

An extensive amount of data is collected while this process takes place, fully describing the participant's interaction with the computer during each photograph, as described in the remaining of the paper.

2 Behavioural Biometrics and Affective Computing

Behavioural Biometrics is the field of study related to measuring uniquely identifying patterns in human actions. Many of the things we do are done in a unique way: walking, talking, typing on a keyboard or using the mouse [4]. Given the uniqueness of our behaviours, Behavioural Biometrics are often used to implement continuous and post-login methods of security and authentication. Nonetheless, they can also be used to detect significant changes in the user's behaviour, which may indicate a significant change in the user's state. In previous

work we have studied the effects of mental fatigue and stress on keystroke and mouse dynamics [5, 6].

In this work, the goal is to determine if different emotions (with variations in their valence, arousal and dominance) can have similar effects on the individual’s interaction with the computer. Ultimately, this will lead to the development of software and devices that are aware of the individual’s emotional state, allowing them to react accordingly. Other researchers have also addressed this topic, albeit considering different features.

Shukla et al. look at the user’s typing behavior to identify emotional states [7]. The authors use a total of 8 features: session time, keystroke latency, dwell time, sequence, typing speed, frequency of error, pause rate and capitalization rate. Questionnaires were used to assess the emotional state of the participants. All this data was then used to train classifiers for human emotion recognition from the keyboard typing patterns. In a related approach, the authors of [8] analyze typing behavior against positive/negative emotions. Its main conclusion is that all participants have shown significant differences in typing patterns when under positive and negative emotions, elicited through facial feedback [9].

3 Architecture

The architecture of the system consists of three main components. The *User Area* is where the participant’s interaction with the system takes place. It includes a desktop with a mouse and keyboard, and three accelerometers attached to the keyboard. The developed application runs in this desktop, showing the photographs to the participant and collecting the data. This data is sent, in real time (whenever the participant advances to the following photograph), to the server.

The *Server Area* contains a Mongo database where data from the several studies are stored. The database also stores the structure of each research protocol created by researchers through the administration dashboard (e.g. sequences of photographs and other settings). All data is stored in the form of JSON documents.

Finally, the *Data Analysis* component provides a set of tools to automate data analysis and facilitate the work of the researcher. This includes data aggregation (e.g. by emotional valence), data extraction (e.g. to raw .csv files), feature extraction, statistical analysis (e.g. hypothesis testing) and training and validation of classifiers for emotional valence.

3.1 Administration

The administration dashboard allows the researcher to manage all the relevant data stored in the database. Specifically, through the dashboard, the researcher or team of researchers associated to a project are able to:

- Manage users - researchers can add/edit/remove participants as well as consult their information (e.g. date of birth, occupation, name);

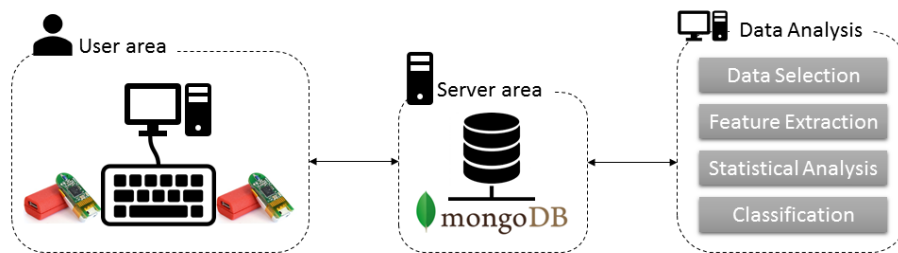


Fig. 1. Main components of the developed system.

- Manage research protocols - researchers can add/edit/remove research protocols, that are constituted by sequences of IAPS photograph identifiers, as well as meta-data (e.g. date of creation, responsible researcher, description);
- Manage participations - researchers can assign participants to specific research protocols. The dashboard allows to visualize individuals who have not yet participated in a given protocol;
- Manage collected data - researchers can visualize a summary of the data collected in a given participation (of a participant in a research protocol). Through the dashboard the researcher can also delete this data.

3.2 Data Collection

Different types of data are generated and collected by the system. Namely, we make a distinction between behavioural data and operational data.

Behavioural data includes data collected from the keyboard and the accelerometers attached to it.

Accelerometers are used with the aim to measure the intensity of typing (increased intensity results in larger values of acceleration). For this end, three WAX3 accelerometers are attached to the back of each keyboard: one in the center and the other two closer to the each side of the keyboard

The WAX3 combines a triple axis accelerometer sensor ($\pm 16g$, $4mg$ resolution) with an ultra low power IEEE 802.15.4 2.4 GHz band. The device also integrates a USB 2.0 enabled microcontroller and micro USB connector along with an on-board power source (rechargeable Li-Polymer battery). This enables the device to behave as a receiver or transmitter and facilitates easy charging and reconfiguration using the bootloader application. Additional functionality can be added using the available expansion port and open source code structure.

Each instance of data generated by each accelerometer contains the following elements: the timestamp, the identification of the accelerometer, and three values denoting the acceleration measured in each of the axes.

While pressure-sensitive keyboards could be used for this purpose, this solution is less expensive and can be applied on any keyboard.

The keyboard, on the other hand, is used by the participant to type the words that describe each image. While the words typed can be useful to understand

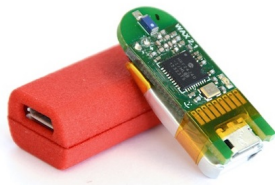


Fig. 2. WAX3 triple axis accelerometer.

the participant’s emotional assessment of each image (through text analysis), the goal is, at the moment, different: we aim to analyse *how* the participant types when visualizing images portraying different emotions.

To this end, the system registers the events related to pressing or releasing keys while typing, together with the key pressed and the timestamp in which it happened. The generated data thus describe two different types of events:

- KEY_DOWN, timestamp, key
Identifies a given key from the keyboard being pressed down, at a given time;
- KEY_UP, timestamp, key
Describes the release of a given key from the keyboard, in a given time.

Finally, operational data describes the actions of the participant throughout the experimental session. This includes the moment in which he/she is shown a new photograph, the moment in which writing starts or the moment in which the photograph is hidden.

Since the emotional rating of each photograph is known, operational data provides the necessary context to interpret behavioural data. For instance, it allows to aggregate and study the participant’s behaviour by emotional valence or arousal, eventually pointing out if different levels of valence or arousal consistently result in specific and different interaction patterns.

The features extracted from these sources of data are described in Section 4.

3.3 Data Storage

The data collected are stored, in real-time, in a Document-oriented Database. These kind of data stores are designed to store and manage documents. Typically, these documents are encoded in standard data exchange (e.g. XML, JSON, YAML, or BSON). These kind of stores allow nested documents or lists as values as well as scalar values, and the attribute names are dynamically defined for each document at runtime.

In the specific case of this work, MongoDB is being used. MongoDB is a database that is half way between relational and non-relational systems. It provides indexes on collections, it is lockless and provides a query mechanism. MongoDB provides atomic operations on fields like relational systems, supports automatic sharding by distributing the load across many nodes with automatic

failover and load balancing. Data are stored in a binary JSON-like format called BSON that supports boolean, integer, float, date, string and binary types. The communication is made over a socket connection.

MongoDB is actually more than a data storage engine, as it also provides native data processing tools: MapReduce and the Aggregation pipeline. Both the aggregation pipeline and mapreduce can operate on a sharded collection (partitioned over many machines, horizontal scaling). These are powerful tools for performing analytics and statistical analysis in real-time, which is useful for ad-hoc querying, pre-aggregated reports, and more. MongoDB provides a rich set of aggregation operations that process data records and return computed results. A significant amount of the features is extracted using either MapReduce or the Aggregation pipeline, as detailed in Section 4.

4 Feature Extraction

One of the key innovative aspects of the proposed environment is the extraction of new features to characterize the effects of emotional valence, arousal and dominance on Human-Computer Interaction. Hence, this section focuses on describing these new features, that will significantly enhance this kind of studies when compared to existing approaches.

Features have been organized in two main groups: operational and behavioural. In this section, the term *dimension* denotes one of the three affective dimensions of each photograph: valence (ranging from pleasant to unpleasant), arousal (ranging from calm to excited) and dominance (ranging from controlled to in-control). Indeed, since the aim is to assess interaction according to emotion, all the features described are calculated by participant and by dimension.

For each of the features described below, the following measures are provided (by participant/dimension): raw values, average, median, standard deviation and quartiles.

4.1 Operational Features

Operational features are those extracted from operational data, that describe the actions of the participant throughout the study. The following features are considered:

- Words/characters - expresses the average number of words or characters that the participant types in a given dimension. Perhaps people tend to be more expressive when experiencing pleasant emotions?;
- Time to start typing - quantifies the time spent by the participant visualizing the photograph before starting to type;
- Time typing - quantifies the time spent typing by the participant in each dimension;
- Total time - quantifies the time spent by the participant in a given photograph (time since first visualizing the photograph until advancing to the next one);

A significant number of other features can be considered in the future, namely those related to Natural Language Processing including sentiment analysis. However, these features are at the moment outside the scope of our work.

4.2 Behavioural Features

Behavioural features, on the other hand, are features that describe the manner in which the participant interacts with the computer, during the study. The following features can be extracted from the system:

- Time between keystrokes - describes the time span between two consecutive KEY_UP and KEY_DOWN events, i.e., how long did it take the user to press another key after releasing the previous one;
- Keystroke latency - quantifies the time span between two consecutive KEY_DOWN and KEY_UP events, i.e., for how long did the user press a given key;
- Typing speed - quantifies how fast a participant types, in words per minutes, considering a word as any group of five characters;
- Keystroke force - describes the acceleration measured on the keyboard, in each of the three accelerometers used and in each of the axes.

4.3 Other Features

Additional features are also available, namely through the MapReduce feature and the Aggregation pipeline of MongoDB. Specifically, we consider features derived from the previously mentioned ones by grouping data by participants' characteristics. These characteristics include age, gender, level of scholarship or occupation. These features may be very interesting, namely to understand how these characteristics affect Human-Computer Interaction or if emotions affect users with different profiles differently.

5 Conclusions

In this paper we described an environment for supporting the carrying out of studies focused on human emotion identification. In traditional approaches, researchers rely on a pencil-and-paper approach, generally using the Self-Assessment Manikin (SAM)[10]. The main disadvantages of this approach is that it is susceptible to error (when the researcher inputs the participant's answers into the computer) and depends on the researcher being present.

The significant contributions of this work are two-fold. One the one hand, the proposed system completely automatizes the whole process, making it distributed and structured. Data collection takes place in real-time and results are available immediately after the conclusion of the protocol. Moreover, researchers can easily share and reuse protocols or studies. It also makes it easier for people other than the researchers to apply the instrument, with the same validity.

On the other hand, this system provides new features that are not available in other approaches and may reveal interesting and new insights. Moreover, these

new features allow the study of the influence of emotion on Human-Computer Interaction, eventually contributing to the further understanding of one of the most interesting topics in this field: that of understanding and reacting to user emotion.

Concluding, we believe that this system can be of interest for the researcher communities studying Human emotion (notably Psychologists), Human-Computer Interaction and Affective Computing.

Acknowledgement

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

References

1. Pinheiro, A., Liu, T., Zhao, Z., Nestor, P.G., McCarley, R.W., Niznikiewicz, M.A.: Emotional cues during simultaneous face and voice processing: Electrophysiological insights. (2012)
2. Cambria, E.: Affective computing and sentiment analysis. *IEEE Intelligent Systems* **31**(2) (2016) 102–107
3. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (iaps): Affective ratings of pictures and instruction manual. Technical report A-8 (2008)
4. Pimenta, A., Carneiro, D., Novais, P., Neves, J.: Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In: *International Conference on Hybrid Artificial Intelligence Systems*, Springer (2013) 222–231
5. Rodrigues, M., Gonçalves, S., Carneiro, D., Novais, P., Fdez-Riverola, F.: Keystrokes and clicks: Measuring stress on e-learning students. In: *Management Intelligent Systems*. Springer (2013) 119–126
6. Carneiro, D., Novais, P.: Quantifying the effects of external factors on individual performance. *Future Generation Computer Systems* **66** (2017) 171–186
7. Shukla, P., Solanki, R.: Web based keystroke dynamics application for identifying emotional state. vol **2** (2011) 4489–4493
8. Tsui, W.H., Lee, P., Hsiao, T.C.: The effect of emotion on keystroke: an experimental study using facial feedback hypothesis. In: *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, IEEE* (2013) 2870–2873
9. Lewis, M.B.: Exploring the positive and negative implications of facial feedback. *Emotion* **12**(4) (2012) 852
10. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* **25**(1) (1994) 49–59