Emotion Recognition in the Manual Interaction with Graphical User Interfaces

Björn Schuller, Gerhard Rigoll, and Manfred Lang
Institute for Human-Machine Communication
Technische Universität München
D-80290 München, Germany
(schuller | rigoll | lang)@ei.tum.de

ABSTRACT

In this paper we introduce a novel approach to human emotion recognition based on manual computer interaction. The presented methods rely on conventional graphical input devices: Firstly a standard mouse as used on desktop PCs, and secondly the interaction on touch-screens or -pads as in public information terminals, palm-top devices or tablet PCs is considered. Additionally the gain of the integration of touch pressure information is evaluated. Four discrete emotional states are classified: irritation, annoyance, reflectiveness, and neutral affect for the use in initiative tutoring, error clarification, Internet customer personalization, and others. The optimal feature-set is discussed and ranked according to a linear discriminant analysis. A working system using Support Vector Machines for the classification is tested in real-life scenarios. The performance of up to 83.2% correct assignment clearly indicates that user emotion recognition is possible without special hardware in any standard graphical user environment independent of the underlying application.

1. Introduction

It is believed that human-computer interaction systems capable of sensing and responding to a user's affective feedback are likely to be perceived as more natural, efficacious, persuasive, and finally more trustworthy [1]. Therefore an advanced HCI faces the challenge of the ability to recognize a user's emotion. A high number of potential application scenarios as tutoring, medical use, high-risk environments, video games, and others exist. Traditional advances propagate a user observation by invasive means such as measuring the heart rate, skin-conductivity or humidity. More recent non-invasive methods on the other hand seem to provide more comfort for the user. A great interest can be observed especially in speech and mimic based emotion recognition [2]. However, all these approaches require special, and sometimes expensive hardware as biosensors, microphones or cameras. Very often the measurement of the input feature stream itself is susceptible to external influences as in vision or acoustic based data capturing. Considering these aspects we introduce emotion recognition based on tactile interaction information with a conventional mouse, touch-screen or touch-pad. The measurement of the input data stream, namely the mouse or touching fingertip movements, can be considered as very robust, and the sensors are present in almost any computer working place, palm-top, tablet PC or information terminal in public places. In the works of Kirsch and Picard [3] or Ark et al. [4] the mouse is also considered a natural place to sense emotional signals form a user's hand. However, in these approaches special hardware is required introducing a mouse that senses pressure patterns. Their feature collection takes only place when the user directly points the mouse at a feedback icon. In our novel method the user's movement itself is analyzed independently of the underlying application and without the need of additional bio or pressure sensors. We decided for four discrete emotion labels, which proved reasonably derivable out of the manual interaction, namely irritation, annoyance, reflectiveness, and a neutral state for the discrimination of emotion presence. Our intended use of the affect information comprises active aid provision only when a user is actually irritated or error clarification when observing an annoyed user subsequent to a system action. A further special interest exists in internet customer guidance where the emotional information includes a double experienced value added effect as the user benefits from personalized interaction and the provider gathers additional valuable information useful for supplementary services. In the following chapters the features used in the recognition of emotion in the mouse- or touch-interaction will be presented.

Subsequently the classification by use of Support Vector Machines, and the acquisition of an emotional corpus and user acceptance in a real-life test scenario will be discussed. Before the final conclusions are drawn a chapter presents the results and discusses these in view of the different hardware provisions.

2. Mouse interaction features

Initially we discuss features used in the recognition of emotion out of the mouse interaction. Later on we show the differences in touch emotion recognition. In our experiment conventional Logitech optical 2-button wheel mice were used at a screen resolution of 1280 x 1024 pixels. In order to build a reasonable model of features providing information of the underlying affect a number of attributes has been considered. We distinguish between two different interaction forms: The off-click movement, meaning no mouse-button is pressed, and respectively the on-click movement. On-click movements often tend to be shorter, as the user just clicks on a button. However, if the user moves an object or selects a menu-item such on-click movements can be of longer total duration as well. We analyze geometrical and temporal aspects of a movement. In a former work [5] we applied a dynamic geometrical contour plus their first and second order derivatives, and Hidden-Markov-Models as a classifier in the recognition. Only touch-events on a touch-screen have been considered at that time. However, our related works in speech emotion recognition [6] showed better performance for derived static features out of the low-level dynamic contours. With respect to these results we conveniently derive high-level features out of the geometrical and temporal contours for each movement. These comprise:

- Total sum of the contour values
- Number of zero-crossings
- Maxima, minima
- Means of the absolute values
- Standard deviations, variances

The same features are also calculated out of the

- Auto-correlation function of the contour
- First order contour derivative
- Second order contour derivative

obtaining a high-dimensional feature vector for each mouse event. As for the geometrical characteristics we construct the ideal line of the shortest connection between the start- and endpoint of a mouse-movement. Next we calculate the shortest distance in pixels for each data-point to the ideal line. As a result we achieve a local deviation contour, which can be interpreted as a rotation of the original mouse-movement contour. Next we derive higher-level attributes as the geometrical length of the ideal line and the above mentioned features. Only the minimum of the deviation contour itself is neglected, as it always equals zero considering the fact that at least the start- and endpoint resemble the ideal connection line. As only locally different data-points have been respected yet, the time delta between two recorded coordinates has been neglected so far. In a second contour we therefore consider this time delay between two locally adjacent points. The derived feature-types of the delay contour resemble the ones of the local deviation contour. Additionally we integrate derived features of the distribution function and its first order derivative.

3. Touch interaction features

The principles of the mouse interaction based emotion recognition can be applied accordingly for touch-interfaces. We decided for a touch-screen that also provides a z-value to be able to evaluate the gain of an additional pressure measurement. However, in our opinion the principles presented can be used as well in pen-based interaction as on hand-held devices, tablet PCs or graphical pads. In our experiments an Elo TouchSystems Intellitouch 2500s Touch-screen equipped with an acoustic surface wave technology iTouch Touch-on-Tube was used. The screen resolution had to be reduced to 800 x 600 pixels due to technical limitations. A 5 MHz-signal pulse is sent to transducers that form the waves on the screen surface. The waves disperse with a speed of 3m/msec. By localization of the wave absorption spot the x and y coordinates can be determined. The pressure is measured by the degree of wave absorption and coded in 8 bits allowing for 256 different values. As a user presses harder the bearing area of his fingertip will increase and therefore absorb a higher amount of wave. Considering the overall extracted features we decided for the same as in mouse interaction concerning the planar Cartesian coordinates. Additionally the z-coordinate and a transform into spherical coordinates lead to further equivalently derived features. In total a 220 dimensional feature vector is used. As one of the main differences to the mouse interaction can be seen in the use of pressure information, the 15 most important feature-ranks can be seen in the following table, where z abbreviates the values of the z-contour, δz the first derivative, $\delta\delta z$ the

second derivative and *acf(x)* the auto correlation function of a contour *x*. The given feature weights are achieved by a Linear Discriminant Analysis (LDA), and are presented normalized to the highest occurring performance.

Table 1. LDA-Ranking pressure based features

Rank	Contour	Type	Weight
1	acf(δz)	Min, max	1.0000
2	acf(δδz)	Min, max	0.9536
3	z	Number of values	0.9360
4	acf(z)	Min, max	0.8944
5	acf(δδz)	Variance	0.8799
6	z	Mean	0.8763
7	δδz	Standard deviation	0.8615
8	δz	Standard deviation	0.8206
9	acf(z)	Number of negative values	0.7428
10	acf(δδz)	Number of negative values	0.7377
11	acf(z)	Standard deviation	0.7276
12	δz	Min, max	0.7265
13	acf(z)	Variance	0.7145
14	z	Variance	0.7136
15	δz	Zero-crossings	0.5913

4. Support Vector classification

Due to their high generalization and discriminant training capabilities Support Vector Machines (SVM) became a popular classification method currently. A basic precondition is linear separability of two classes. Otherwise a transform into a higher dimensional feature space is used, where linear separation is possible. In our case a radial basis function kernel as mapping function showed good performance. In order to achieve discriminant separation of two classes a hyper-plane is placed with maximum distance between the references of the classes by a Lagrange optimization. The so-called support vectors span this plane. As only these vectors are needed as references a great reduction is achieved by this approach. In the recognition phase the distance of a test-sample to the separation hyper plane forms the basis of decision. One SVM was trained for each class against all other classes and the SVM with the minimum distance for the trained class was selected throughout the recognition process.

5. Emotional data acquisition

In order to obtain real emotion data we decided for a playful interrogative background dialog approach. Five users were equipped with a program running on their computers during their everyday work. Three more users used the program while interacting on a touch screen. From time to time the program initiatively asked them about their feelings subsequent to their interaction in order to label their recent mouse or touch interactions. 1983 labeled emotional data samples of mouse interaction and 825 of touch interaction could be collected over a period of six months avoiding anticipation effects and ensuring adequate occurrences of all four affects. In a second phase the system occasionally asked the user if the recognized emotion was correct.

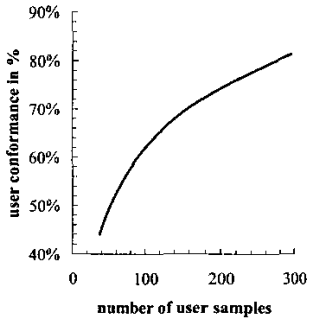

Figure 1. User conformance in the adaptation

This was done in order to get an impression of the conformance with the sheer mathematical recognition results. The system was initialized with 250 exemplary data sets of other users and 50 samples of the actual user. As more data of the user herself was accessible the foreign training samples were gradually replaced resulting in 300 user specific samples and no foreign data sets after around two weeks of adaptation process. The final user conformance exceeded 83% resembling the achieved maximum recognition rates. Figure 1 shows the increasing user conformance with the ongoing adaptation to the end-user.

6. Results and discussion

In this chapter the results obtained on the collected data sets are presented. The model consistency was initially guaranteed by a reclassification. The convincing reclassification result of 99.9% encouraged real test series with training disjoint data. In three cycles one third of the data was used for testing and two-thirds for training. Considering mouse interaction the overall recognition rate was 83.19%. The following tables show the averages of the obtained recognition rates. In this and the following table *ntl* abbreviates neutral, *irr* irritation, *ann* annoyance, and *ref* reflectiveness. To the right hand side the recognized

emotions will be found. The values resemble correct recognition rates in %.

Table 2. Confusion matrix mouse interaction

Emotion	ntl	irr	ann	ref
ntl	**90.2**	2.3	4.8	2.7
irr	19.4	**76.5**	3.5	0.6
ann	9.4	1.2	**88.4**	1.0
ref	14.5	1.3	6.5	**77.7**

The following table shows the performance for touch emotion recognition.

Table 3. Confusion matrix touch interaction

Emotion	ntl	irr	ann	ref
ntl	**58.9**	20.1	8.4	12.6
irr	0.0	**92.7**	7.3	0.0
ann	0.8	13.5	**85.7**	0.0
ref	7.1	33.4	16.2	**43.3**

The overall recognition rate was 70.12%. However, clear differences in the maximum performance can be observed between different emotions. Annoyance is more easily recognized in both approaches, while reflectiveness shows less good results. In general the decrease in performance using touch information is partly due to the 173% higher resolution used in mouse interaction compared to the touch screen. Anyways the touch coordinates cannot be determined as exactly as the mouse coordinates, as the fingertip possesses a large bearing area resulting in inaccurate localization of the coordinates and movements. One further main difference between mouse and touch interaction is the fact that no off-click movements exist on a touch-screen, as the touch itself is interpreted as click. Therefore only sparse longer movements are observed. This fact makes it harder to interpret the emotion in a touch and shows that the off-click information is more important in the estimation. This might differ considering pen-based input, where hand-written input allows for interpretation of longer data sequences. However, in that case it might be difficult to isolate the emotion information from the written content. Furthermore a different behavior can be observed when positioning the touch screen horizontally or vertically due to ergonomic factors. In the vertical position physical weaknesses of the user may falsify the results due to a trembling arm. If the emotions are further restricted to three or two labels, performance rates increase significantly. Providing a final comparison of feature importance, it can be said that features derived of the spherical coordinates showed greater potential than the Cartesian based features. The temporal aspects showed in both cases less contribution than the geometrical information. The least gain however was obtained by the addition of the z-information. This indicates that a comparable performance can be achieved on any touch-panel providing only information of the movement on a parallel plane to the touch surface.

7. Conclusion

We believe that the results of this novel approach to user emotion recognition clearly demonstrate that there is emotion in mouse and touch interaction. Even without special hardware and independent of the underlying application the affect of a user could be recognized with up to 83.2% correct assignment providing more comfort for the user compared to invasive bio measurements and avoiding extra costs for sensors. Considering these promising results everyday devices and software could be enhanced with the capability to recognize and react to user emotions. In future works multi-touch capable devices, and new affects will be investigated with more test-users. Finally long-term tests need to be performed in order to analyze the influences of physical conditions or response to the content provided by the computer.

8. References

[1] M. Pantic, L. Rothenkrantz: "Toward an Affect-Sensitive Multimodal Human-Computer Interaction," *Proc. of the IEEE Vol. 91*, pp. 1370-1390, Sep. 2003.

[2] R. Cowie et al.: "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine Vol. 18*, no. 1, pp. 32–80, Jan. 2001.

[3] R.W. Picard: "Toward computers that recognize and respond to user emotion," *IBM Systems Journal Vol. 39*, NOS 3&4, pp. 705-719, 2000.

[4] W. Ark, D. C. Dryer, and D. J. Lu: "The Emotion Mouse," *Proc. of HCI International '99*, Munich, Germany, August 1999.

[5] B. Schuller, M. Lang, G. Rigoll: "Multimodal Emotion Recognition in Audiovisual Communication", *Proc. of the ICME 2002*, CD-Rom Proceedings, Lausanne, Switzerland, 2002.

[6] B. Schuller, G. Rigoll, M. Lang: "Hidden Markov Model-Based Speech Emotion Recognition," *Proc. of the ICASSP 2003 Vol. II*, pp. 1-4, Hong Kong, China, 2003.