Smile Detection for User Interfaces

O. Deniz, M. Castrillon, J. Lorenzo, L. Anton, and G. Bueno

First and last authors: Universidad de Castilla-La Mancha, E.T.S.I.I
Campus Universitario, Avda. Camilo Jose Cela s/n, 13071, Spain
Oscar.Deniz@uclm.es
Rest of the authors: Universidad de Las Palmas de Gran Canaria
Dpto. Informatica y Sistemas,Campus de Tafira, Edificio de Informatica
35017 Las Palmas, Spain

Abstract. Perceptual User Interfaces (PUIs) aim at facilitating human-computer interaction with the aid of human-like capacities (computer vision, speech recognition, etc.). In PUIs, the human face is a central element, since it conveys not only identity but also other important information, particularly with respect to the user's mood or emotional state. This paper describes both a face detector and a smile detector for PUIs. Both are suitable for real-time interaction. The face detector provides eye, mouth and nose locations in frontal or nearly-frontal poses, whereas the smile detector is able to give a smile intensity measure. Experiments confirm that they are competitive with respect to extant detectors. These two detectors are used in an unobtrusive application that allows to interact with an Instant Messaging (IM) client.

1 Introduction

The words that we speak account for only a part of the meaning that we convey. Tone, body language and facial expression communicate the rest. Perceptual User Interfaces use multiple input modalities to capitalize on all the communication cues, thus maximizing the bandwidth of communication between a user and a computer. The human face is the main source of information for short-distance interaction. Thus, much computer vision research is being devoted to face perception. Face detection, for example, has nowadays become a basic task in many perceptual interfaces. The number of face detection systems proposed in the literature is significant, see for example [1,2]. Still, a number of challenges remain in terms of real-time performance, ability to extract facial features, non-frontal face detection, etc.

The ability to show and interpret emotions is crucial for human interaction. Detecting and modeling user's emotions can therefore be considered another goal of Perceptual User Interfaces. In this respect, the human smile is a distinct facial configuration (suggesting that it may not be very difficult to detect) and can be very informative. Smile detection can be used in any application that requires to assess the user's state such as distance learning systems, patient monitoring, film ratings, etc.

G. Bebis et al. (Eds.): ISVC 2008, Part II, LNCS 5359, pp. 602–611, 2008.

Instant Messaging, on its part, is a form of real-time communication based on typed text. Since IM appeared in the 1970s to facilitate communication with other users logged in to Unix machines, it has expanded enormously. Currently, IM is actively used as a fast communication tool, specially among young people and in the workplace. A number of enhancements and capabilities have been added in the last years. Despite advances in clients and network speeds, however, current IM software is still based on typed text. The well-known *emoticons* are used as an attempt to convey user's facial expression or emotion. The lack of verbal and visual cues can otherwise cause what were intended to be humorous, sarcastic, ironic, or otherwise non-100%-serious comments to be misinterpreted, resulting in arguments. Nevertheless, the user has to specifically type the keystroke sequence of the emoticon to show. User status (i.e. online, away, etc.) also has to be specifically controlled by the user. User status is not a trivial aspect of IM communication. A typical misunderstanding occurs when someone is writing to you but you forgot to change your status to 'Away'. The other user may interpret that you were simply ignoring him/her.

In this work it is shown how face and smile detection algorithms are applied to enhance user experience with IM software, thus taking advantage of the widespread availability of webcams, particularly in modern laptops. The paper is organized as follows. Sections 2 and 3 describe the face detector and the smile detector, respectively. Experiments are shown in Section 4. The application for IM software is described in Section 5. Finally, the main conclusions and future lines of research are outlined.

2 Face Detection

The face detection system used for this work (see [3]) integrates, among other cues, different classifiers based on the general object detection framework by Viola and Jones [4], skin color, multilevel tracking, etc.

The Viola-Jones object detector is a cascade of classifiers. Each classifier uses a set of Haar-like features. The classifiers are 'weak': each one has a very high detection ratio, with a small true reject ratio. This way they act as a filter chain. Only those image regions that manage to pass through all the stages of the detector are considered as containing a face, see [5]. For a cascade of K classifiers, the resulting detection rate, D, and the false positive rate, F, of the cascade are given by the combination of each single stage classifier rates:

$$D = \prod_{i=1}^{K} d_i \qquad\qquad F = \prod_{i=1}^{K} f_i \tag{1}$$

On the other hand, this framework allows a high image processing rate, due to the fact that background regions of the image are quickly discarded while spending more time on promising face-like regions.

In order to further minimize the influence of false alarms, the facial feature detector capabilities were extended, locating not only faces but also eyes, nose

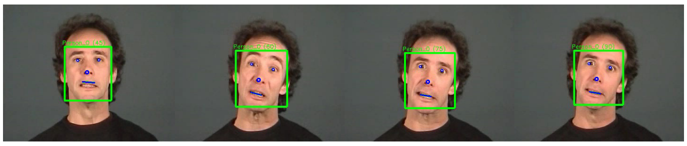


Fig. 1. Facial element detection samples for a sequence extracted from DaFEx [7]

and mouth. This reduces the number of false alarms, for it is less probable that multiple detectors, i.e. face and its inner features, are activated simultaneously with a false alarm.

Positive samples for the training sets of inner features were obtained by annotating manually the eye, nose and the mouth locations in 7000 facial images taken randomly from the Internet. The images were later normalized by means of eye information to 59×65 pixels. Five different detectors were computed: 1-2) Left and right eye (18×12 pixels), 3) eye pair (22×5), 4) nose (22×15), and 5) mouth (22×15). These detectors have been made publicly available, see [6].

The facial element detection procedure is only applied in those areas which bear evidence of containing a face. This is true for regions in the current frame, where a face has been detected, or in areas with a detected face in the previous frame. For video stream processing, given the estimated area for each inner feature, candidates are searched in those areas not only by means of Viola-Jones' based facial features detectors, but also by SSD-tracking previous facial elements. Once all the candidates have been obtained, a likelihood based on the normalized positions for nose and mouth is computed for each combination, selecting the one with the highest likelyhood. Fig. 1 shows the possibilities of the described approach with a sequence extracted from DaFEx [7].

3 Smile Detection

The new Sony Cybershot DSC T-200 digital camera has an ingenious "smile shutter" mode. Using proprietary algorithms, the camera automatically detects the smiling face and closes the shutter. To detect the different degrees of smiles by the subject, smile detection sensitivity can be set to high, medium or low. Some reviews argue that: *"the technology is not still so much sensitive that it can capture minor facial changes. Your facial expression has to change considerably for the camera to realize that"*[8], or *"The camera's smile detection – which is one of its more novel features – is reported to be inaccurate and touchy"*[9]. Whatever the case, detection rates or details of the algorithm are not available, and so it is difficult to compare the system. Canon also has a similar smile detection system.

Sensing component company Omron has recently developed a "smile measurement software", which measures the amount of happiness that human subject of a photo are exhibiting [10]. The software uses a proprietary 3D model fitting technique to detect and analyze faces. This smile checking software rates how much a subject is smiling and gives a 'smile factor' on a scale of 0 to 100%.

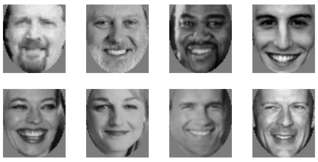


Fig. 2. Examples of positive images used for training the smile detector

This analysis only takes about 44 milliseconds using a PIV at 3.2Ghz and can be performed on images of faces as small as 60 pixels wide. Omron claims that this device is more than 90% accurate.

On a more scientific level, there are a significant number of papers that have tackled facial expression recognition, see the surveys [11,12]. Few systems, however, have been specifically designed for smile detection. The smile detector of [13] used a vector of lip measures (extracted from an edge image) and a perceptron classifier. Edge features, however, may no be robust enough for practical use. More elaborated is the method of [14], which used HLAC (Higher-order Local Autocorrelation) along with Fisher weight maps, achieving recognition rates of 97.9%. The BROAFERENCE system was developed to assess TV or multimedia content through smile measurement [15]. In this case, 8 mouth points are tracked, feeding a neural network classifier with the 16 feature vector. Unfortunately the authors do not give precise figures for its performance, although they claim that it achieves a 90% detection rate [16].

The smile detection system proposed in this paper is based on a Viola-Jones cascade classifier. Training was carried out using 2436 positive images and 3376 negative images. The images were first extracted from Internet, then detected and normalized by the face detection system described above. Figure 2 shows some examples of the positive images used for training.

When the cascade detector is searching over the image, it may produce multiple positives around the positive region (the smile). Those detected rectangles largely overlap. Usually, isolated detections are false detections and they should be discarded. The number of neighbor detections is normally used as a confidence threshold.

For smile detection, the number of neighbor detections can also be considered as a confidence measure. The more neighbors detected around an image region, the more confidence that the region contains a smile. If the negative images of the training set contain mostly neutral faces then the number of neighbors can be considered as a measure of smile intensity. Figure 3 shows some of the faces used in the negative set.

4 Experiments

4.1 Face Detection

The face detection system was tested with 74 video sequences corresponding to different individuals, cameras and environments, with a resolution of 320x240.

Fig. 3. Examples of negative images used for training the smile detector

They represent a single individual sat and speaking in front of the camera or moderating a TV news program. The face pose is mainly frontal, but it is not controlled, i.e. lateral views and occlusions due to arm movements are possible. The eyes are not always visible. The total set contains 26338 images.

In order to test the detectors performance, the sequences were manually annotated, therefore the face containers are available for the whole set of images. However, eye locations are available only for a subset of 4059 images. The eyes location allows us to compute the actual distance between them, which will be referred below as *EyeDist*. This value will be used to estimate the goodness of eye detection. Mouth and nose detection were not analyzed.

Two different criteria have been defined to establish whether a detection is correct: a) Correct face criterium: A face is considered correctly detected, if the detected face overlaps at least 80% of the annotated area, and the area difference is not doubled, and b) Correct eye criterium: The eyes of a face detected are considered correctly detected if for both eyes the distance to manually marked eyes is lower than a threshold that depends on the actual distance between the eyes, *EyeDist*.

Table 1 shows the results obtained after processing the whole set of sequences with different detectors. The correct detection ratios (TD) are given considering the whole sequence, and the false detection ratios (FD) are related to the total number of detections. As for the face detector, it is observed that it performs more than twice faster than Viola-Jones' detector. Speed was the main goal in our application, the face detector is critical for the intended application.

4.2 Smile Detection

In order to test smile detection, experiments were carried out using a set of 4928 images of 108 individuals. The images were previously processed by the

Table 1. Results for face and eye detection processing using a PIV at 2.2Ghz. More details in [3].

	Viola-Jones [4]		Face detector used here [3]	
	TD	FD	TD	FD
Faces	97.69%	8.25%	99.92%	8.07%
Left Eye	0.0%	-	91.83%	4.04%
Right Eye	0.0%	-	92.48%	3.33%
Proc. time	117.5 msecs.		45.6 msecs.	

Table 2. Comparison of smile detectors

Parameter	This paper	Omron [10]	Shinohara & Otsu [14]	Ito et al [13]
Training images	5812	?	72	1800
Test images	4928		24	3-min video
Individuals (test set)	108		4	3
Detection rate	96.1% (with 16 stages)	more than 90%?	?	97.5%
False acceptance rate (FAR)	2.3% (with 16 stages)	?	?	18%
Recognition rate (over the two classes)	98.3% (at 16% FAR)	?	97.9%	?
Processing time per image	45.6ms on a PIV at 2.2Ghz) + 0.36ms (on a Core2 Duo at 2.4Ghz)	44ms on a PIV at 3.2Ghz	less than 50ms on a PIV at 1.8Ghz	?

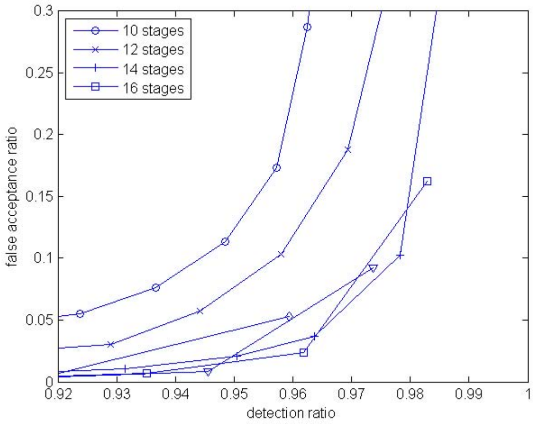


Fig. 4. ROC curves for the smile detection system. With more than 16 stages the detection rate may be considered too low to be useful, they were not shown in order to keep the Figure uncluttered.

face detector, see similar examples in Figures 2 and 3. This particular set and the individuals were different from those used for training. Figure 4 shows the ROC curve for smile detection. Detection rates are above 96% with less than 3% false acceptance rate. This would compare well with Omron's system, of which we only know that is more than 90% accurate. On the other hand, the smile detector spends on average 0.36ms per (normalized) face image (running on a Core^{TM}2 Duo CPU at 2.4Ghz, using a 16 stage classifier). This means that the total (face detection+smile detection) processing time per image is roughly 46ms. Table 2 shows a comparison with other three smile detection systems.

The ability to estimate smile intensity was also put to test. In this case, a different dataset was used. In the already mentioned DaFEx database 8 professional actors showed 7 expressions (6 basic facial expressions + 1 neutral) on 3 intensity levels (low, medium, high). The 'happy' pictures were extracted of the database sequences (see Figure 5), and the intensity level was compared with

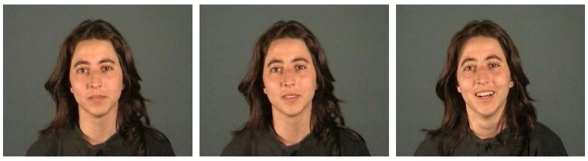


Fig. 5. Low, medium and high intensity happy expressions from DaFEx

Table 3. Average number of neighbors obtained for the low, medium and high intensity smiles, over a total of 3440 images

Smile intensity	Number of neighbors	
	Mean	Std. dev.
Low	7.63	5.31
Medium	11.31	8.01
High	18.40	10.09

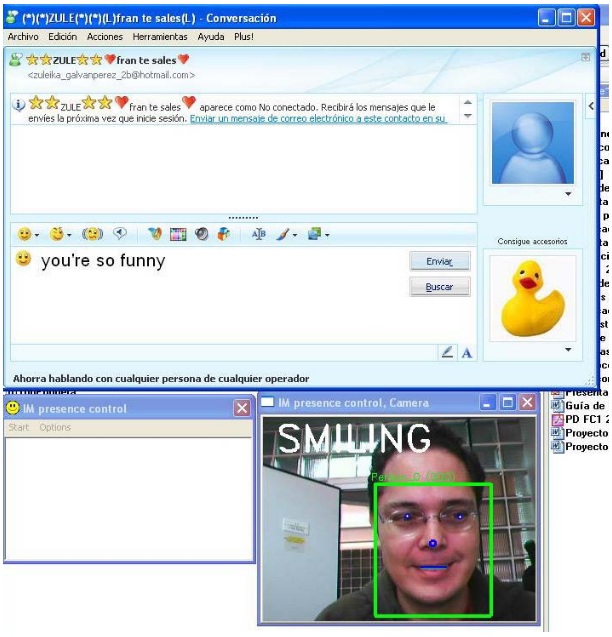
the number of neighbors given by the smile detection system. Table 3 shows the results. It can be seen from the table that smile intensity (as given by the database labels) and the number of neighbors are correlated.

Still, as the intensity discretization is sparse in the DaFEx database (i.e. only low, medium and high labels), a second database was tested. The Japanese Female Facial Expression (JAFFE) [17] database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image had been rated on 6 emotion adjectives by 60 Japanese subjects (on 5-level scale, 5=high, 1=low). This database allowed us to have numerical intensity values for the happy emotion (the averages of the scores given by the 60 subjects). The correlation ratio between these values and the number of neighbors given by the smile detection system was 0.64 (95% confidence interval: $[0.55, .., 0.71]$). Again, this supports the fact that the number of neighbors is a good indicator of smile intensity.

5 Application: Instant Messaging Presence Control

The face and smile detection systems described above were used in an application aiming at enhancing IM communication. The application uses a standard webcam to measure both presence (user status) and smile. In particular, it can control two features of the IM client: Away/online status and Smile emoticons.

The application developed is able to detect when the user is in front of the laptop or away. The smile detector automatically inserts smile emoticons in the conversation window when the user is smiling. High intensity smiles can also be detected, using the number of neighbors as a measure. The IM client application is controlled through keystrokes sent to its window (as specified by its title). Keystrokes sent will not interfere with user's typing.


(*)(*)ZULE(*)(*)(L)fran te sales(L) - Conversación
Archivo Edición Acciones Herramientas Ayuda Plus!
ZULE fran te sales
<zuleika_galvanperez_2b@hotmail.com>
Consigue accesorios
Enviar
Buscar
IM presence control
Start Options
IM presence control, Camera
SMILING
Guía de
PD FC1
Proyecto
Proyecto

ZULE fran te sales aparece como No conectado. Recibirá los mensajes que le envíes la próxima vez que inicie sesión. Enviar un mensaje de correo electrónico a este contacto en su

you're so funny

Ahorra hablando con cualquier persona de cualquier operador

Fig. 6. IM Presence Control

Keystroke strings, both of status change and of the emoticons to insert, can be adjusted by the user. An example keystroke string is " :-) ", the typical smiley. Special keys can be inserted too, between "¡" and "¿". For example, the string "¡HOME¿ ¡HOME¿)¡HOME¿-¡HOME¿:¡END¿" would insert a smiley (plus a blank space) at the beginning of the current text line in the conversation window (the ¡HOME¿ key must be sent before each character because in IM clients the conversation window is continually placing the cursor at the end of the line). Status change can be typically achieved with strings such as "¡ALT¿ade" that navigate through the options of the main menu.

The options of the application include: IM application window title string, IM conversation window title string, Smile keystroke string, Big smile keystroke string (typically ":-D"), Time between sending of smile keystroke strings (in seconds, 0 to wait for a no-smile before sending a new smiley), Away keystroke string, Online keystroke string, Time without face before sending an Away keystroke string (in seconds), Smile detections before a smile or big smile keystroke string is sent, Sensitivity (the smaller the more smile detections), Smile/Big smile threshold and Show/hide live video window (the video window is hidden by default).

The application can be executed with the argument '-s', which makes it start automatically and remain minimized in the tray. This way it will run unobtrusively. In Figure 6 the live video window is shown, when working with Windows Live Messenger (Copyright of Microsoft Corp.).

6 Conclusions

Perceptual User Interfaces aim at facilitating human-computer interaction with the aid of human-like abilities like computer vision, speech recognition, etc. In PUIs, the human face is a central element, since it conveys important information, particularly with respect to the user's mood or emotional state. This work proposes both a face detector and a smile detector for PUIs. Both can work together in real-time with modern commodity hardware. The face detector provides eye, mouth and nose locations in some situations, whereas the smile detector is able to give a smile intensity measure. Experiments confirmed that they are competitive with respect to extant detectors. These two detectors have been used in an unobtrusive application that allows to control the user status of an Instant Messaging (IM) client. The application can also automatically insert smile/big smile emoticons in the IM client conversation window. As far as the authors know, it is the first time that such computer-vision-based aid is added to IM communication.

Future work shall include the use of the smile detector in other applications that could take advantage of joy assessments: film previews, email clients, intelligent desktops, human-robot interaction, video games, wearable computing, etc. The natural extension would be to use the same methods described here to build a general facial expression recognizer which can give intensity values. Another aspect for future work is the effect of other parts of the face other than the mouth. Smiles can involve subtle cheek raising around the eyes (the so-called Duchenne smile). However, this may not be a reliable cue, not least because it does not appear in every smile.

Acknowledgments

This work was partially supported by Project ULPGC07-008 of Universidad de Las Palmas de Gran Canaria.

References

1. Hjelmas, E., Low, B.K.: Face detection: A survey. Computer Vision and Image Understanding 83, 236–274 (2001)
2. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. Transactions on Pattern Analysis and Machine Intelligence 24, 34–58 (2002)
3. Castrillón, M., Déniz, O., Hernández, M., Guerra, C.: ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. Journal of Visual Communication and Image Representation, 130–140 (2007)
4. Viola, P., Jones, M.: Robust real-time face detection. IJCV 57, 151–173 (2004)
5. Hewitt, R.: Seeing with OpenCV, part 2. Servo, 48–52 (2007)
6. Reimondo, A.: OpenCV Swiki. Last accessed: January 2008 (2008)
7. Battocchi, A., Pianesi, F.: Dafex: Un database di espressioni facciali dinamiche. In: SLI-GSCP Workshop Comunicazione Parlata e Manifestazione delle Emozioni (2004)

8. Entertainment.millionface.com: Smile detection technology in camera. Last accessed: January 2008 (2008)
9. Swik.net: Sony's Cyber-shot T200 gets its first review. Last accessed: January 2008 (2008)
10. Omron Corp.: Omron OKAO vision system. Last accessed: January 2008 (2008)
11. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. Pattern Recognition 36, 259–275 (2003)
12. Pantic, M., Member, S., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 1424–1445 (2000)
13. Ito, A., Wang, X., Suzuki, M., Makino, S.: Smile and laughter recognition using speech processing and face recognition from conversation video. In: Procs. of the 2005 IEEE Int. Conf. on Cyberworlds, CW 2005 (2005)
14. Shinohara, Y., Otsu, N.: Facial expression recognition using Fisher weight maps. In: Procs. of the IEEE Int. Conf. on AFGR (2004)
15. Kowalik, U., Aoki, T., Yasuda, H.: Broaference - a next generation multimedia terminal providing direct feedback on audience's satisfaction level. In: INTERACT, pp. 974–977 (2005)
16. Kowalik, U., Aoki, T., Yasuda, H.: Using automatic facial expression classification for contents indexing based on the emotional component. In: EUC, pp. 519–528 (2006)
17. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Procs. of the Third IEEE International Conference on Automatic Face and Gesture Recognition (1998)