AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit

Daniel McDuff
Affectiva
Waltham, MA, USA
daniel.mcduff@affectiva.com

Abdelrahman Mahmoud
Affectiva
Waltham, MA, USA
mahmoud@affectiva.com

Mohammad Mavadati
Affectiva
Waltham, MA, USA
mohammad.mavadati
@affectiva.com

May Amr
Affectiva
Waltham, MA, USA
may.amr@affectiva.com

Jay Turcot
Affectiva
Waltham, MA, USA
jay.turcot@affectiva.com

Rana el Kaliouby
Affectiva
Waltham, MA, USA
kaliouby@affectiva.com

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2890247

Abstract
We present a real-time facial expression recognition toolkit that can automatically code the expressions of multiple people simultaneously. The toolkit is available across major mobile and desktop platforms (Android, iOS, Windows). The system is trained on the world's largest dataset of facial expressions and has been optimized to operate on mobile devices and with very few false detections. The toolkit offers the potential for the design of novel interfaces that respond to users' emotional states based on their facial expressions. We present a demonstration application that provides real-time visualization of the expressions captured by the camera.

Author Keywords
Facial expressions; Emotion; Affective Computing

ACM Classification Keywords
H5.2. User Interface

Introduction
Emotions play an important role in everyday life and human-computer interaction can benefit greatly from emotion sensing [1,5,7,9]. Systems that respond to the emotions of a subject can be made to appear more

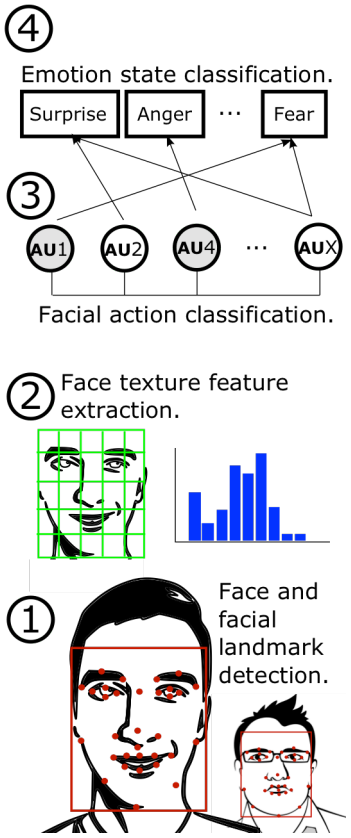

Figure 1: Automated facial coding pipeline. 1) Detection of face(s) and localization of the key facial landmarks on each face. 2) Extraction of texture features using HOG. 3) Classification of facial actions. 4) Modeling of prototypic emotions using EMFACS.

empathetic and provide a much more natural user interface.

The face is one of the richest channels of expression. It communicates both emotions and social signals. The Facial Action Coding System (FACS) [3] is the most comprehensive and widely used objective taxonomy for coding facial behavior. Facial action units (AUs) are the building blocks of facial expressions. Manual coding of FACS is extremely laborious and time consuming. Furthermore, it is impractical for any real-time or scaled application. In the past, systems for automated coding of facial expressions have been constrained due to the limited availability of training data. Using a web-based framework, similar to [6], we collected videos of hundreds of thousands of individuals. These videos were coded by expert FACS coders to provide a rich dataset of facial expression examples. We trained state-of-the-art facial action and emotion classifiers using this data. In this paper we present the AFFDEX software development kit (SDK). The SDK provides an easy interface for processing multiple faces within a video or live stream in real-time. The SDK has cross-platform capabilities.

Automated Facial Coding

Our system for automated facial coding has four main components: 1) Face and facial landmark detection, 2) face texture feature extraction, 3) facial action classification and 4) emotion expression modelling. Figure 1 shows an overview.

Face and Facial Landmark Detection

Face detection is performed using the Viola-Jones face detection algorithm [10]. Landmark detection is then applied to each facial bounding box and 34 landmarks identified. If the confidence of the landmark detection is below a threshold then the bounding box is ignored. The facial landmarks, head pose and intraocular distance for each face are exposed in the SDK.

Facial Actions

Histogram of Oriented Gradient (HOG) features [2] are extracted from the image region of interest defined by the facial landmark points. Support Vector Machine (SVM) classifiers, trained on 10,000s of manually coded facial images collected from around the world, are used to provide scores from 0 to 100 for each facial action. For details of the training and testing scheme see [8].

Emotion Expressions

The emotion expressions (Anger, Disgust, Fear, Joy, Sadness, Surprise and Contempt) are based on combinations of facial actions. This coding was built on the EMFACS [4] emotional facial action coding system. The emotion expressions are given a similar score from 0 (absent) to 100 (present).

Gender and Glasses

In addition to facial action and emotion expression classifiers the SDK has classifiers for determining gender and whether the person is wearing glasses.

The classifiers have two operating modes: *static* and *causal*. The static classifiers allow classification of single images. The causal classifiers leverage temporal information available in video sequences to further increase the accuracy of the facial expression measures.

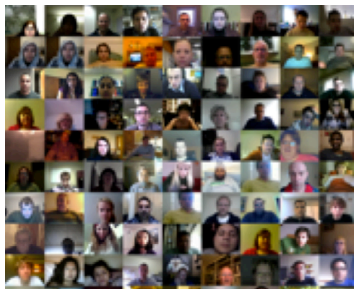

Figure 1: Example images taken from the world's largest dataset of facial videos from over 75 countries in very diverse lighting and surroundings.


Disgust


Fear


Joy


Sadness


Surprise


Anger


Contempt

Table 1: The emotion expressions that are detected by the AFFDEX SDK. Each emotion is given a score from 0 (absent) to 100 (present).

Performance Evaluation

The system was tested on an independent set of 10,000 images to verify the generalizability of algorithms. We did not control the lighting or pose of the participants. Figure 2 shows example frames.

AFFDEX SDK

We have created a Software Development Kit (SDK) to allow easy integration of the software into other applications. The hardware available impacts the number of frames that can be processed per second. Typically it is possible to achieve frame rates of 10 Frames per Second (FPS) on mobile devices and 30 FPS on laptop/desktop devices. Our demonstration will allow the users to get real-time feedback on their facial expressions. Figure 3 shows an example of the mobile interface for our demo. Participants will be able to try interacting with an iPad and Android tablet. Figure 4 shows the desktop demo that will allow users to test the multi-face capability on a larger screen. The SDK demo application code and the SDK are available to download at: http://www.affectiva.com/*sdk*

Applications

Emotion sensing offers a vast amount of potential for improving human-computer and human-human interactions. Below are some of the emerging applications that the SDK supports:

Video Conferencing. Non-verbal cues are critical to effective communication. In remote interactions these cues are lost. Systems that could recover affective signals would make remote communication easier.

Automated Tutors/Online Education. As distance learning becomes more popular automated measurement of learners' emotional states becomes more critical. Having access to affective data could help educators improve the quality of content.

Life-logging and Health. Emotions impact many aspects of our daily lives. However, we often have difficulty reflecting on our emotional states. Life-logging tools and devices that help track emotions would be very useful [5,7,9] and would provide a way of linking lifestyle patterns to changes in mood.

Gaming. Computer games that respond to human emotions offer a new dimension to game-play with characters that adapt naturally to the player.

AU1 In. Brow Raise	**AU2** Out. Brow Raise	**AU4** Brow Furrow	**AU9** Nose Wrinkle	**AU10** Upper Lip Raise	**AU12** Lip Corner Pull	**AU15** Lip Depress	**AU17** Chin Raise	**AU18** Lip Pucker	**AU20** Lip Press	**AU25** Mouth Open	**AU28** Lip Suck	**AU43** Eyes Closed	**Smirk***

Table 2: The facial actions that are detected by the SDK. Each action is given a score from 0 (absent) to 100 (present). *Smirk is defined as an asymmetric lip corner pull (either on the right or left side of the face but not both)

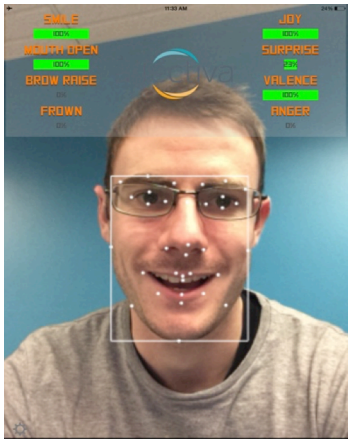

Figure 3: Screenshot of the iOS SDK demo application.

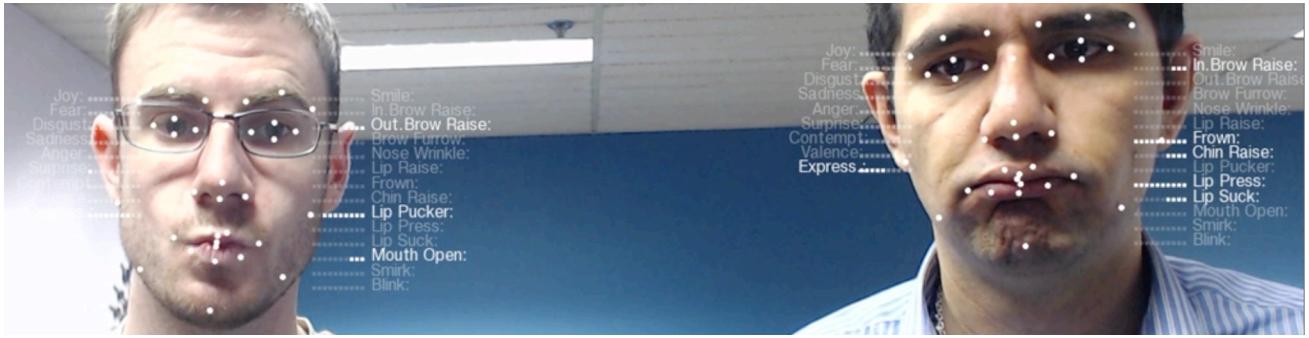

Figure 4: Screenshot of the real-time multi-face expression classification SDK demo application.

Conclusion

We believe that the availability of emotion sensing software will have a significant impact on the design of connected devices and interfaces. Our SDK provides state-of-the-art real-time multi-face expression analysis for multiple platforms.

References

1. Brave, S., and Nass, C. 2003. "Emotion in human–computer interaction." *Human-Computer Interaction*: 53.
2. Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.
3. Ekman, P. and Friesen, W. 1978. "Facial Action Coding System: A technique for the measurement of facial movements." *Consulting Psychologist*.
4. Friesen, W. and Ekman, P. 1983. "EMFACS-7: Emotional facial action coding system." *Unpublished manuscript, University of California at San Francisco* 2: 36.
5. McDuff, D., Karlson, A., Kapoor, A., Roseway, A., and Czerwinski, M. 2012. "AffectAura: an intelligent system for emotional memory." In *Proc of CHI '12.*
6. McDuff, D., Kaliouby, R. and Picard, R. 2012. "Crowdsourcing facial responses to online videos." *Affective Computing, IEEE Transactions on* 3.4: 456-468.
7. Picard, R., *Affective computing*. MIT press, 2000.
8. Senechal, T., McDuff, D. and Kaliouby, R., 2015. "Facial Action Unit Detection Using Active Learning and an Efficient Non-Linear Kernel Approximation." In *Proc. ICCV IEEE `15.*
9. Ståhl, A., Höök, K., Svensson, M., Taylor, A., and Combetto, M. 2009. "Experiencing the affective diary." *PERS UBIQUIT COMPUT,* 13.5: 365-378.
10. Viola, P. and Jones, M. 2001. "Rapid object detection using a boosted cascade of simple features." PROC CVPR IEEE '01.