
A Facial Affect Mapping Engine

Leonardo Impett

Department of Engineering
University of Cambridge
li222@cam.ac.uk

Tadas Baltrušaitis

Computer Laboratory
University of Cambridge
Tadas.Baltrusaitis@cl.cam.ac.uk

Peter Robinson

Computer Laboratory
University of Cambridge
Peter.Robinson@cl.cam.ac.uk

Abstract

Facial expressions play a crucial role in human interaction. Interactive digital games can help teaching people to both express and recognise them. Such interactive games can benefit from the ability to alter user expressions dynamically and in real-time. In this demonstration, we present the Facial Affect Mapping Engine (FAME), a framework for mapping and manipulating facial expressions across images and video streams. Our system is fully automatic runs in real-time and does not require any specialist hardware. FAME presents new possibilities for the designers of intelligent interactive digital games.

Author Keywords

Augmented Reality; Facial Affect; Face Tracking; Facial Puppetry; Face Swapping; Intelligent Games.

ACM Classification Keywords

I.3.7 [Three-Dimensional Graphics and Realism]

Introduction

In this demonstration, we present the Facial Affect Mapping Engine (FAME) – a framework for modifying and transferring facial expression and facial identity from images/videos of one face (the animator) to another (the avatar).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
IUI'14 Companion, February 24–27, 2014, Haifa, Israel.
ACM 978-1-4503-2729-9/14/02.
<http://dx.doi.org/10.1145/2559184.2559203>

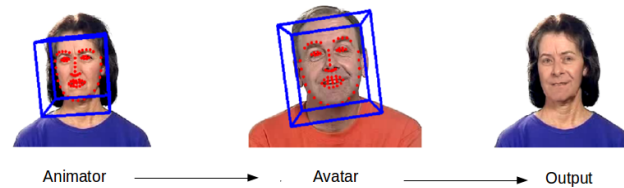


Figure 1: The role of animator (controlling expression) and avatar (controlling identity) in the system. The user can be either or both of these

FAME has three modes of operation that can be described as an interaction between a user, captured by web-cam, and a reference source image or video:

- (a) A user live-animates a face from a source image. The reference face (avatar) is then blended onto the user's video-stream. Figure 1 (with the user playing the role of an animator) and Figure 2 show examples of this mode.
- (b) The source video acts as the animator and the user as the avatar. Thus the user's face is shown displaying the facial expression in the source video. This might be used to teach a child with an autism spectrum condition how to express emotions.
- (c) The user's facial expression and head pose are altered and transferred back to the user's video stream (see Figure 3 for an example). This can be used to magnify or attenuate the intensity of an emotional expression.

Related work

A fundamental part of FAME is facial puppetry – where a puppeteer animates a digital avatar using their own face. Similar to our work, Saragih *et al.* [6] created a facial puppetry system that runs in real-time and requires only a single unlabelled avatar image in neutral expression.

A similar facial puppetry system has been used for a psychological experiment that explored the effect on attenuation of head-pose and facial expression in photo-realistic videoconferencing scenarios [3]. Our system enables researches to conduct similar experiments, with natural and realistic blending of the modified face onto the original video background.

Face swapping is also of relevance to this system, and research on this has tended to focus on image-based manipulation [2]. Successful systems for replacing faces in single-camera video have been developed [4, 7]. However, such approaches either rely on computationally expensive 3D models or specialist 3D scanning hardware [8].

Architecture

The Facial Affect Mapping Engine uses techniques from both facial puppetry and face swapping to transfer identity and expression through video streams. Our system, unlike existing systems, uses three stages: face tracking, expression manipulation, and video re-synthesis.

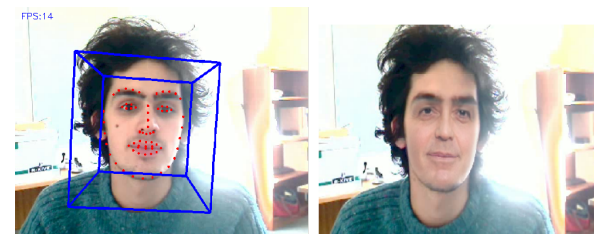


Figure 2: Example of the system in action, demonstrating the facial landmark tracking and identity transfer

The user and reference videos can switch roles as 'puppet/avatar' or 'puppeteer/animator'. Face tracking operates on both puppet and puppeteer in parallel, whilst expression manipulation combines the two, along with any

expression modification. The output video is synthesised by re-situating the animated puppet in either the user or reference video streams.

In order to track the geometry of the face, we need to track landmark points on the face together with the head pose. For tracking faces we use the fully automatic Constrained Local Neural Field model [1]. The tracked points are shown in Figure 2.

To manipulate expressions, we extract the faces from both images, modify the expression in the puppeteer's face and map that expression onto the puppet's face. Having tracked the important feature points in the puppet's and puppeteer's video streams, we extract the tracked face image from the whole image frame by cropping. We then triangulate the face images with the points tracked, and piecewise affine-warp these faces to a neutral shape. This leads to the expression in a video source being separated into a texture (image) and shape (triangular mesh). Changes in expression influence both texture and shape, but these can now be controlled separately.

We can attenuate or exaggerate facial expression, we can also change rotation and translation of the head (see Figure 3).



Figure 3: Attenuation of expression using the expression vector. Left to right: increasing expression amplitude

Facial expression and head pose both produce changes in the face texture, due to changes in illumination, wrinkles and so on. For realism these textural changes must also be present in the animated face. Such dynamic features from the puppeteer texture must be combined with the static features (identity) from the puppet texture. For this we used a video expression ratio image (ERI), based on the approach outlined by Liu *et al.* [5].

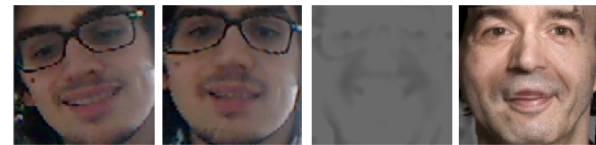


Figure 4: Our four-stage ERI system (left to right): the original puppeteer image; the puppeteer image warped to a standard shape; the resulting ERI map; ERI texture transferred onto the puppet image

To resynthesise the face, the expression avatar texture is warped to the modified shape of the puppeteer. Hardware-assisted graphics is used to blend the new face and the background image.

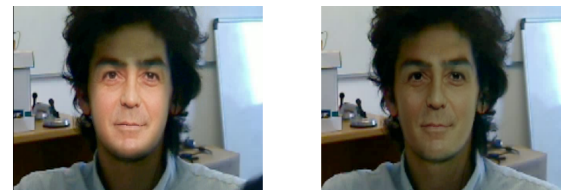


Figure 5: An extreme case of face re-colouring: left, the animated blended directly onto background. Right, the face re-coloured using our approach, then blended on

The output from ERI has been normalised in luminance, and inherits its colour distribution directly from the

puppet's initial image. It therefore needs re-colouring in order to blend well with the puppeteer's video stream. We assume the colour spectra (in RGB space) roughly follow Gaussian distributions, and linearly scale and shift the spectra to equalise the mean and standard deviation for each colour component. This technique is robust to large differences in skin colour and lighting conditions between the puppet and the puppeteer (see Figure 5).

The facial region from the puppeteer image is blurred heavily, to eliminate facial features (which have a higher spatial frequency) but preserve face-scale lighting information (which has a low spatial frequency). The animated puppet face is then blended with this undercoat, which re-lights the puppet's face appropriately for the scene with very little computational overhead, reducing any artifacts caused by the puppeteer's facial features.

Conclusions

We have presented a Facial Affect Mapping Engine, a tool for use in interactive intelligent games where the modification, transposition or generalisation of facial expression of emotion is important.

Both the puppet and puppeteer can be controlled either from a reference video or from a standard web-cam, giving rise to three distinct modes of operation. The synthesised face is graphically blended into either the puppet or puppeteer's source video with a high degree of realism.

The software produces near-photorealistic results in a wide range of cases and in real time (at 15-30 fps on a 2.4 GHz PC with 8GB of RAM).

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework

Programme (FP7/2007-2013) under grant agreement n 289021 (ASC-Inclusion).

References

- [1] Baltrušaitis, T., Robinson, P., and Morency, L.-P. Constrained local neural fields for robust facial landmark detection in the wild. In *ICCV workshops* (2013).
- [2] Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P., and Nayar, S. K. Face swapping: automatically replacing faces in photographs. In *ACM Transactions on Graphics (TOG)*, vol. 27, ACM (2008), 39.
- [3] Boker, S. M., Cohn, J. F., Theobald, B.-J., Matthews, I., Brick, T. R., and Spies, J. R. Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars. *Phil. Trans. Royal Soc.: B* 364 (2009).
- [4] Dale, K., Sunkavalli, K., Johnson, M. K., Vlasic, D., Matusik, W., and Pfister, H. Video face replacement. In *ACM Transactions on Graphics* (2011).
- [5] Liu, Z., Shan, Y., and Zhang, Z. Expressive expression mapping with ratio images. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM (2001), 271–276.
- [6] Saragih, J. M., Lucey, S., and Cohn, J. F. Real-time avatar animation from a single image. In *FG* (2011).
- [7] Tzeng, V., Cheng, Y.-T., Liang, Y., Wang, C.-C., Chen, B.-Y., Chuang, Y.-Y., and Ouhyoung, M. 3d-model-based face replacement in video.
- [8] Weise, T., Li, H., Van Gool, L., and Pauly, M. Face/off: Live facial puppetry. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM (2009), 7–16.