

Segment and Browse: A Strategy for Supporting Human Monitoring of Facial Expression Behaviour

Michael J. Lyons, Mathias Funk, and Kazuhiro Kuwabara

ATR Intelligent Robotics & Communication Laboratories,
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan 619-0288
mlyons@atr.jp

Abstract. We describe a system to ease long-term human monitoring of mood via facial expressions. Video images are processed in real-time to isolate the area of the face and record facial expressions. Optic flow is used to annotate motion of the face. A simple-to-use browser is used to navigate the facial expression record. A preliminary evaluation of both components is reported.

1 Introduction

An increasingly large fraction of the population of many nations is aged and no longer capable of a self-sufficient lifestyle. In Japan such seniors often live with family members and can require many hours of attention each day. Our team is investigating technologies to assist the caregivers of dependent seniors as well as improve the quality of life for all concerned. A sub-goal is to develop methods to aid the assessment of the emotional status of the dependent person. This could reduce the need for continuous monitoring of a dependent, as well as improve efficiency in responding to their needs. Fully automatic systems are under investigation, however it is also very important to also consider partially-automatic systems which augment human propensities to assess another's emotional status and needs. Facial expressions provide salient information on mood [2, 4]. There has been substantial progress on the research problem of classifying facial expressions [3], however automatic systems are not very effective at combining contextual information with facial expression data in judging affect [4]. A major design goal in this project is to create a system which augments human capabilities at judging affect from facial expressions. We automate the long-term observation of facial expressions and format the image data so it can be evaluated quickly by a caregiver. Here we describe a complete working prototype for monitoring and browsing facial expressions over extended periods.

2 System Design and Implementation

The system consists of vision and browser modules. The **vision module**: (a) ambiently acquires images of a person (b) segments the face from the image (c) calculates motion sensitive features (d) stores facial images with motion attribute data. The **browser module** aids a caregiver to navigate the large dataset quickly and assess mood.

2.1 Vision Module

Video input from a camera is reduced from 640x480 pixels to 320x240. A face detection algorithm determines whether or not there is a face in view of the camera and, if so, where it is. The optical flow field of the image is calculated on multiple time scales to detect motion in the video frame. The segmented faces are saved with averages of the optic flow over the entire internal region of the face.

We use a powerful and widely popular algorithm for face detection [5] known as cascade-correlation. The algorithm automatically registers a rectangle of fixed aspect ratio with the head with the face centered in the interior. It detects faces accurately for out of plane rotations of up to roughly 30 degrees and roughly the same amount of in plane rotation of the head. The facial image is scaled to a standard size and stored with timestamp at up to 30 fps. Normalizing the size and position of the face and isolating it from a complex background makes the face data easier to browse.

Optic flow measures displacements of image areas due to motion in the visual scene. For a fixed camera and approximately constant illumination, the optic flow field gauges the local velocities. The method we use for calculating optic flow is known as *block matching* or *block correlation* [1]. With the block correlation technique, a region from one image frame is matched to a region of the same size in the subsequent frame. The velocity vector is calculated from the displacement of the block between frames. To gauge motion at a range of velocities we calculate the optic flow field between frames separated by 1,3,5,7, and 10 video frames. Flow values are

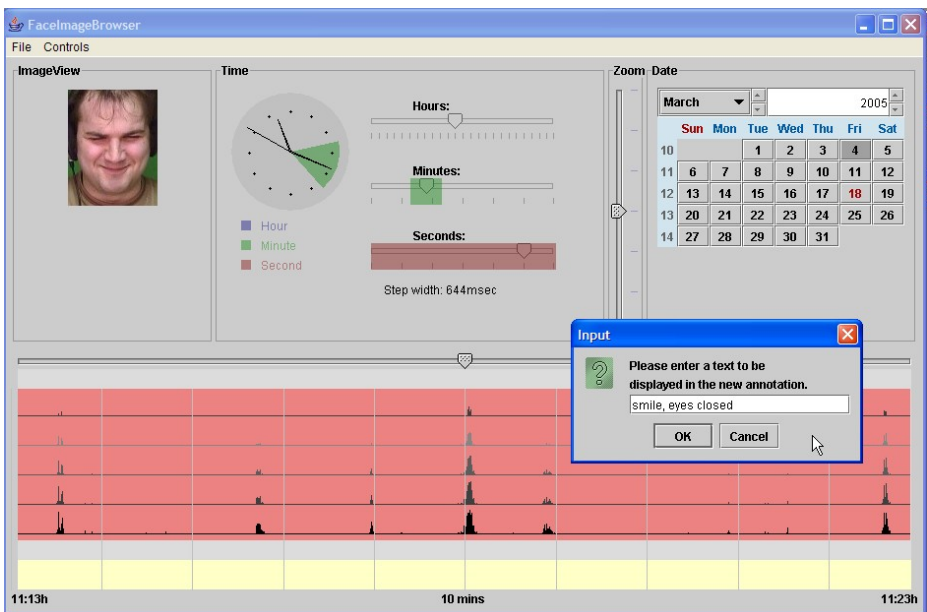


Fig. 1. The facial expression data browser. A detailed description of the components is given in section 2.2.

average over regions corresponding to the internal part of the face as detected by the face detection module. The optic flow values are saved to index file which also contains facial image filenames.

2.2 Browser Module

Figure 1 shows a screen shot of the browser module. The components of this module are (clockwise from upper left corner) (a) normalized display of the facial image at the selected time (b) time range display in clock and linear form (c) a time resolution zoom control to select range of the timeline (d) a calendar for selecting images of a certain date and (from middle to bottom) (e) a timeline for scrolling the time (f) a display of average optic flow value at different velocity scales for each point in time (g) bar indicating the presence of user annotations. A sample annotation window is shown overlapping the optic flow display.

The user selects the date and then zoom the time range control to choose the timeline range. Images are browsed with the timeline scroller. Presence or absence of facial images is indicated by red or grey colour in the optic flow indicator giving an immediate estimate of whether the subject was present. The optic flow annotations provide indication of whether motion of the face occurred. The distribution of activity over the five velocity levels indicate whether the motion corresponds to large rigid body motion of the entire face or more subtle movements corresponding to expressions and speech. Short text annotations may be added and edited or removed by clicking on the lowest active zone below the timeline.

3 Evaluation

We evaluated the capability of the vision module to properly segment facial images in realistic situations. Video of three senior citizens suffering from various stages of dementia was collected in actual domestic situations by our research team. The subject watched television while a roughly frontal view (± 45 degrees) of the upper body and face was recorded. The backgrounds were typically complex. To protect the privacy of these patients, images from this dataset cannot be shown here. The performance of the automatic face detection and segmentation algorithm was found to be more than adequate for our needs, with very few facial views being lost over the course of approximately 30 total minutes recording.

A preliminary user study of the browser module was conducted to test the hypothesis that segmentation of the facial image and annotation with motion data would allow users to navigate facial expression more efficiently. We videotaped 4 subjects while they watched a 20 minute cartoon. Five experimental users had to search these video tapes for target facial expression images selected from the video data. There were 4 target images for each of the 4 videos. We compared average time to complete the task using our browser with that using QuickTime Player and the raw, unsegmented video. Average time to complete a search task using the browser was 58 seconds while the average using QuickTime was 145 seconds. This difference was found to be significant to better than 5% using the Wilcoxin signed ranks test. In addition, accuracy was higher with our browser than with the QuickTime player.

4 Conclusion

We have developed a system to support the monitoring of mood from ambiently captured video data of facial expression behaviour. By segmenting the face from a complex scene, normalizing it, and annotating it with motion data we were able to greatly increase the efficiency of navigation of a large facial expression dataset. In our continuing work on this project we plan to conduct more extensive field and user studies.

Acknowledgement

This work was supported in part by the National Institute of Information and Communications Technology.

References

1. Barron, J. L., Fleet, D. J., Beauchemin, S. S.: Performance of optical flow techniques. *The International Journal of Computer Vision*, 12 (1994) 43-77
2. Ekman, P. , Friesen, W.V.: *Unmasking the Face*. Prentice Hall, Englewood Cliffs (1975)
3. Lyons, M.J., Budynek, J., Akamatsu, S.: Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (1999) 1357-1362
4. Russell, J.A.: Core Affect and the Psychological Construction of Emotion. *Psychological Review* 110 (2003) 145-172
5. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2001) 511-518