An automated video recommendation system to enhance engagement levels in moderate-Dementia care patients

Priya Ganapathy[a], Tejaswi Tamminedi[a], Evan Dong[a], Shalini Keshavamurthy[a], Jacob Yadegar[a]
Aravind Kailas[b], Parminder Juneja[b], Boyd Davis[b] and Dena Shenk[b]

a: UtopiaCompression Corporation
11150 West Olympic Blvd. Suite 820
Los Angeles, CA 90064, USA
{priya, tejaswi, evan, shalini, jacob}@utopiacompression.com

b: University of North Carolina, Charlotte
9201 University City Blvd. | Charlotte, NC 28223-0001
{aravind.kailas, pjuneja, bdavis, dshenk@uncc.edu}

ABSTRACT
We present a non-intrusive system (SENSEI software) that can measure engagement levels in moderate to late stage dementia-care (DCA) patients. The measurement software communicates with the YouTube recommendation system via a web-based program to alter or continue the video clip. The system fuses information from a webcam (face recognition, body posture, and voice intonation features) and body-worn sensors (heart rate and galvanic skin conductance) to determine the arousal and valence levels of patients watching these videos. The proposed technology is aimed at augmenting DCA patient's interest levels that are strong indicators of quality-of-life enhancements.

Categories and Subject Descriptors
I.5.4 [**Pattern Recognition**]: Applications – *Computer vision, Signal processing, Recommendation algorithms, Dementia care.*
General Terms: Algorithms, Measurement, Design, Experimentation.
Keywords: Improving engagement in Dementia-care patients, Emotion Recognition; Multi-modal Probabilistic Information Fusion; Facial Expression Analysis; Voice Intonation Analysis; Body Posture Analysis; Physiological monitoring;

1. SETUP
The demonstration setup illustrated in Figure 1 consists of a personal computer or an internet-enabled TV that streams YouTube videos and a set of non-invasive sensors including video cameras, acoustic sensors, etc. A suite of commonly used wearable sensors (like a Q sensor watch [better to give a reference here]) for measuring physiological cues (such as heart rate, galvanic skin response, and skin temperatures) are also included.

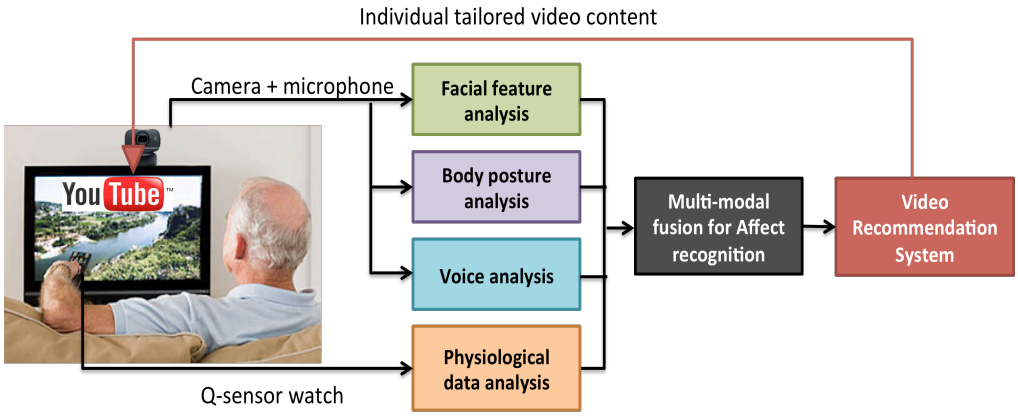

Figure 1. Our proposed automated content recommendation system

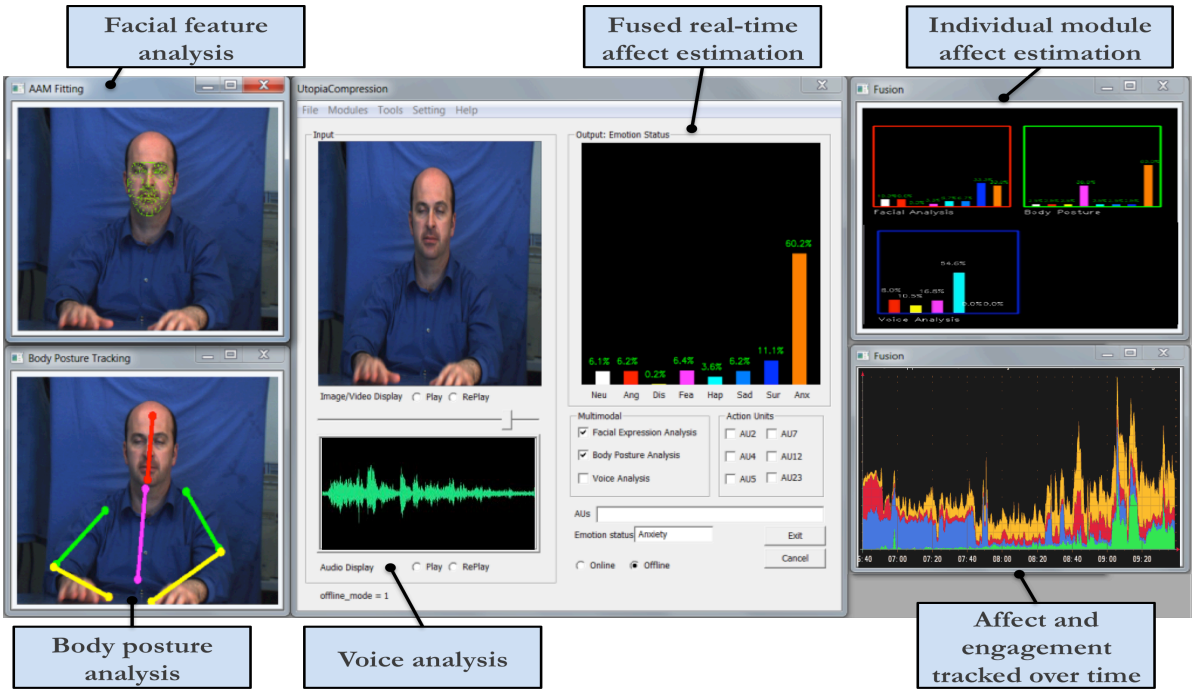

Figure 2. User interface that demonstrates the output of SENSEI software. Demo video obtained from the FABO database with permission [5].

2. DEMONSTRATION DESCRIPTION

As shown in Figure 2 (an intuitive graphical user interface of the SENSEI software), we will demonstrate five (4) major components/ capabilities of our multi-modal engagement measurement/recognition system: (A) Facial expression based emotion recognition module: analyzes the facial muscle state and implements eye tracking to infer the affect and engagement level of the human subject. In this module, robust real-time face detection/ tracking algorithm will first be applied to detect regions of interest (ROIs) (face/head) [1]. A large number of active appearance model (AAM)-based facial features will then be computed on detected ROIs [2]. Finally, those computed facial features will be fed into a machine learning classifier (such as SVM) for relevant emotion recognition. (B) Body posture analysis module: estimates the human body parts' positions to help determine the subject's emotional state based on the well-known pictorial structure representation of the human body. Our body posture analysis module is based on a Bayesian

Network-based upper body tracking system that incorporates various constraints on body parts such as their symmetry, relative size, motion constraints, etc., to name a few, to realize a physically and anatomically feasible offline modeling of the body [3]. In this way, this module can eliminate many false postures, reduce the posture search space and improve the robustness of the body posture detection and tracking. (C) Voice analysis-based affect recognition module: overcomes the prevelant shortcomings in voice/speech feature selection and noise modeling by creating a huge set of features over noisy, real data and then efficiently trimming the features down to only the most relevant sets without losing valuable information. The voice recognition module also computes both local and global features as well as larger prosody cues, indicating both short-term and long-term emotional status of a human subject. (D) Analysis of heart rate and galvanic skin response (GSR): has been implemented using traditional feature extraction algorithms to classify the arousal state based on physiological signals. (E) Multi-modality data fusion module: integrates the aforementioned modalities into a probabilistic data fusion module that is based on the Influence Diagram theory [4] to comprehensively estimate a subject's emotional states based on the available sources of information.

The affect information (arousal and valence) provided by the SENSEI software while watching a video is transformed into a decision criteria for liking a video, disliking a video and/or adding the video as a favorite. This calibration process was established by conducting a small pilot study at UtopiaCompression Corporation (UC) on normal volunteers to watch several clips of YouTube videos and allows the user to tag the videos as 'like,' 'dislike' and an option to add them to their favorites. The trend of valence and arousal data collected while subjects are watching the videos was correlated with their decision to like/dislike a video. We applied traditional machine learning algorithms on the results of this pilot study [6]. The classifiers once trained can classify a video into like/dislike category based on the input valence and arousal data from subject viewing the video. The classifier which performed best on the training/test data was used as the candidate classifier.

Upon the SENSEI software calibration, we developed a web program (using YouTube API) that uses SENSEI software output and classifier to automatically annotate the viewed YouTube video. The design and testing of the program to communicate with YouTube to stream or alter videos in a seamless fashion will be performed prior to deploying the unit in dementia patient rooms. In collaboration with University of North Carolina, Charlotte (UNCC) data will be collected on Dementia care patients. We will re-calibrate the automated annotations because dementia subjects may elicit either low or very high response compared to normal volunteers for the same videos. The SENSEI software will be tweaked to fit the geriatric population (face features, body posture will change with age). This will involve UNCC and UC experts to watch changes in subject's facial features, body posture, eye tracking, and physiology data while they were viewing videos and readjust the annotations. The classifier will be re-trained based on the adjusted ground truth.

3. CONCLUSION

We will demonstrate a multi-modal system that consists of a set of feature extraction and data analysis modules for heterogeneous sensory data and a novel probabilistic information fusion model to accurately estimate the interest levels in subjects watching short YouTube video clips. Based on the continuous evaluation the video content can be altered to sustain engagement in subjects. Our SENSEI software has multiple applications ranging from persistent and non-intrusive monitoring of mentally ill patients in their home environment to development of intelligent tutoring systems to increase interest in students.

4. ACKNOWLEDGMENTS

The work presented in this demonstration is supported under a U.S. DARPA SBIR program (Contract No. N10PC20172). The authors would like to thank U.S. DARPA for the financial support.

5. DISCLAIMER

The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

6. REFERENCES

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. Of IEEE International conference on Computer Vision and Pattern Recognition (CVPR)*, pp. I-511- I-518, 2001.

[2] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6): 681-685, 2001.

[3] L. Zhang, J. Chen, Z. Zeng and Q. Ji, "A generic framework for 2D and 3D upper body tracking, machine learning for human motion analysis: Theory and Practice," *IGI Global*, 133-151, 2009.

[4] Uffe B. Kjaerulff, Madsen, and L. Anders, "Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis," *Proc. Information Science and Statistics,* 2008.

[5] H. Gunes and M. Piccardi, "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior," *Proc.18th IAPR International Conference on Pattern Recognition (ICPR)*, vol. 1, pp. 1148-1153, 2006.

[6] R. Duda, E. Hart, and G. Stork. Pattern Classification (2nd Edition), Wiley, New York. 2001.