Facial Recognition: An Enabling Technology for Augmented Cognition Applications

Denise Nicholson[1], Christine Podilchuk[2], and Kathleen Bartlett[1]

[1] DSCI, Human Systems Engineering, Orlando, Florida
[2] DSCI, STAR Technologies, Eatontown, New Jersey
{Dnicholson,cpodilchuk,kbartlett}@dsci.com

Abstract. Research in Augmented Cognition (AugCog) investigates computational methods, technologies, and non-invasive neurophysiological tools to adapt computational systems to the changing cognitive state of human operators to improve task performance. Closed-loop AugCog systems contain four components: 1) operational or simulated environment, 2) automated sensors to monitor and assess cognitive state via behavior and/or physiology, 3) adaptive interface, and 4) computational decision architecture that directs AugCog adaptations. Since cognitive state is influenced by environment, a critical challenge for AugCog systems is capture of situational awareness (SA) within the decision architecture. Previously, AugCog systems have been demonstrated within simulated environments that provide SA and ground truth data to drive intelligent decision architecture. In live operating environments, electronic C4 systems (i.e., communications), provide a limited model of operator "state," but emerging facial recognition/analysis technology can provide detection, identification, and tracking of humans in the environment to increase the accuracy of the AugCog system's SA.

Keywords: Augmented Cognition, facial recognition, situation awareness, biometrics, environmental monitoring.

1 Augmented Cognition

Building on advances in the fields of neuroscience, cognitive science, and computer science, Augmented Cognition research focuses on the real-time cognitive state of the operator [1]. Current AugCog methods, techniques, and applications range from academic research to industrial/ military operational and training systems to computing and entertainment devices [2]. To enhance human performance, Aug Cog uses physiological and neurophysiological sensors in a closed loop system (see Fig. 1) to detect when the human's cognitive capacity, which fluctuates due to fatigue, stress, overload, or boredom, cannot meet mission demands [3]. Neurophysiological- and physiological-based assessment of cognitive states relies upon a variety of data, including cardiac measures, electroencephalogram (EEG), and functional near-infrared (fNIR) imaging, to evaluate cognitive ability in diverse environments [3]. These real-time, non-invasive measures of an operator's cognitive state can be used to trigger adaptive automation that enhances human performance [2].

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 104–111, 2011.

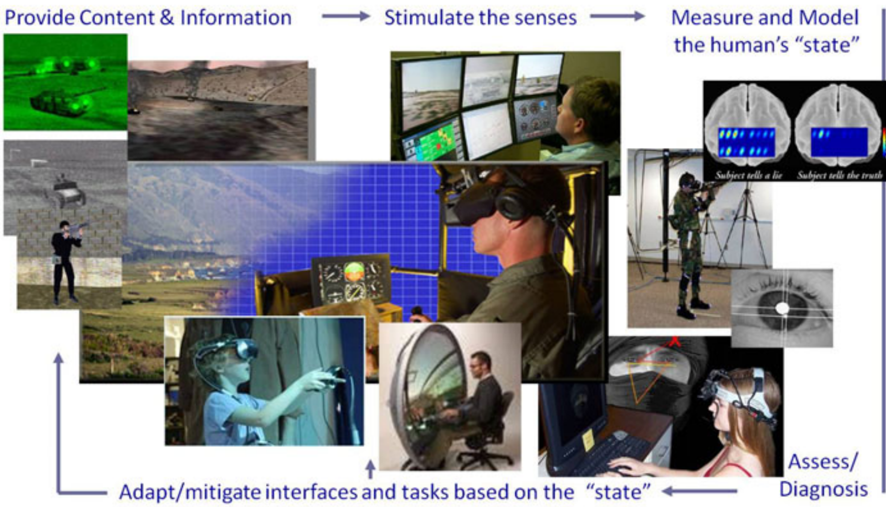

Fig. 1. Model for Adaptive Interaction in Closed Loop System (CLHS) [4]

Regardless of the field of use, the core components of the CLHS, including skill assessments, physiological and behavioral measurements, and mitigations, are extendable concepts. Coupled with field specific task modeling, a flexible system design can identify how users can best achieve and retain mastery of multiple tasks / skills through individualized adaptive interaction. In situations of overload, when the demand for speed or attention exceeds human ability, technology may be activated to compensate for human performance degradation [5] in simulated training and live environments. In tests at a base in Orlando, for example, researchers observed the degraded performance caused by information overload: when soldiers operated a tank while monitoring remote video feeds, they often failed to see nearby targets [6]. Augmented reality (AR) systems, which combine real and artificial stimuli, prepare trainees to operate successfully in a wide range of dangerous, unpredictable environments by monitoring operator state and providing technology-assisted support, via mitigation strategies, when performance overload is indicated. In live situations, mitigation strategies have the potential to save lives.

A mitigation strategy is an intervention technique (i.e., adaptive interaction, operational performance support, task cooperation, and individualized embedded training strategies) driven by the task analysis, automated measures, and diagnostic assessment, that significantly improves human-system performance and enhances skill proficiency and retention. The iterative mitigation process can be mapped to Norman's (1988) "seven stages of action" model, which was developed to represent the human action cycle as people accomplish goals (Fig. 2). Within the adaptive design environment, mitigation strategies are triggered in stages one to three of the "seven stages" model by *perceiving* a user's behavioral or physiological activity through automated measures and diagnostic assessment. Mitigation covers stages four to six [4].

Mitigation strategies can only be effectively triggered if measurement of operator state reflects a complete picture of the operating environment. If the operator registers stress, for example, the digital audio (C4 channels) may reflect voices, but the number and type of individuals in the environment cannot be represented in the AugCog SA in order to trigger the supporting mitigation strategy. Facial recognition technologies may be useful to monitor and report on humans in the environment, thus providing a missing puzzle piece in the monitoring of operator state.

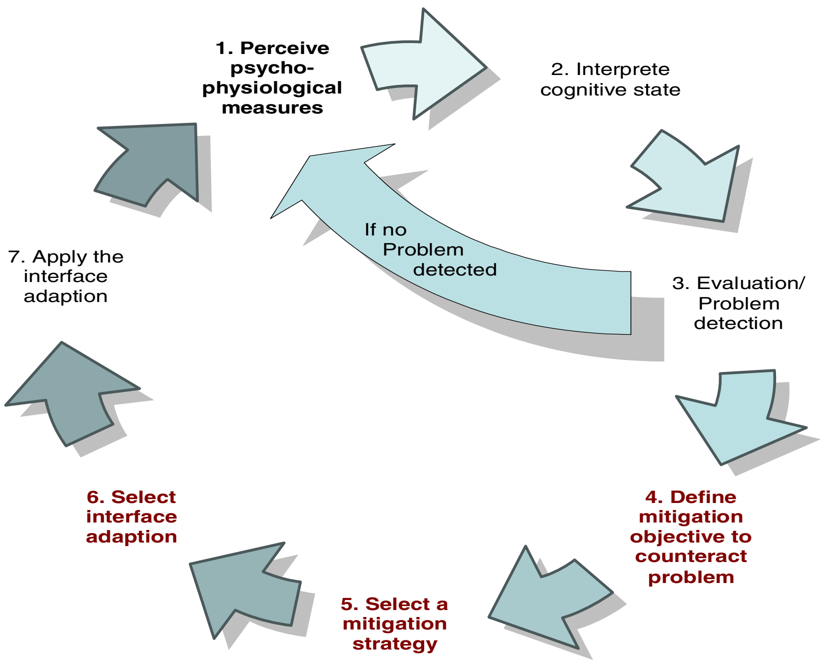

Fig. 2. Norman's "7 Stages of Action" modified for mitigation strategy design [4]

2 Visual Environment and Cognitive State

A complimentary definition of this domain space is provided by Neuroergonomics, which "postulates that the human brain, which implements cognition and is itself shaped by the physical environment, must be examined in interaction with the environment in order to understand fully the interrelationships of cognition, action, and the world of artifacts" [7]. Therefore, collection of comprehensive environmental information is essential when monitoring cognitive state. However, a critical challenge in developing a full closed-loop AugCog system is the accurate representation of the operational environment or situational awareness (SA) within the decision architecture's expert model used to drive system adaptations. Accurate SA data is critical to achieve correct interpretation of the physiological sensors for cognitive state determination of the operator and to ensure appropriate recommendations for system adaptations which optimize human-system performance, under specific environmental conditions.

To address this challenge, most of the previous AugCog closed-loop systems have been demonstrated within fully characterized simulated environments that provide the necessary situational awareness and ground truth data about the operating state to drive the intelligent decision architecture. Other demonstrations have explored live operating environments which provide an opportunity to monitor electronic C4 ISR systems (i.e., communications) to supply the necessary SA. However, the C4 channels don't contain all information influencing the human operator. For example, operators collect additional information about objects, personnel, and targets via their visual field of view, which is not represented in the C4 data. Consequently, a decision system that doesn't have access to the visual information in the operating space is limited in achieving a fully accurate model of the operator's "state."

In any operating environment, the presence (or absence) of other persons represents one of the most fundamental elements of operator state and frequently influences whether or not operator action is required. In a simulated AugCog environment, the entities populating any given situation are programmable, and the manipulation and measurement of operator response can be based on predictable encounters with anticipated threats. In live situations relying on AugCog support, measurement of SA lacks input about the numbers and types of humans in the live environment, although these elements exert significance influence on the cognitive state of the individual. Emerging video surveillance and facial recognition/analysis technology has the potential to provide detection, identification, and tracking of the presence of human targets in the operator's live environment. With this information, diagnosis of the incongruency between the ideal expert state and the actual human state of the operator might result in appropriate adaptations to ensure optimum operating state. Facial recognition technology can be a critical solution for providing the necessary environmental SA for effective diagnosis and driving of AugCog adaptations.

3 Facial Recognition

A suite of facial recognition applications currently being developed at DSCI provides real-time face recognition in uncontrolled environments using novel algorithms for pattern recognition that are robust for differences in facial expression, pose, illumination, camera angle, and facial occlusions. This technology can locate, analyze, and provide information about the faces of personnel located within a live environment and can specify or determine the number of faces in an image. This technology can also be applied to images extracted from a video stream where faces are detected, identified, and tracked through a video clip or across cameras, detecting and responding to occlusions. The approach consists of a general mapping between two images, followed by a measurement of different properties of the two-dimensional vector field representing the mapping task. Since it does not depend on domain specific features, this general approach can be applied to any object or target recognition.

Working in tandem with the target recognition algorithm is a fast search technique for rapid identification via large databases. This fast search algorithm can be applied to any algorithm that computes similarity scores for the purpose of pattern recognition. This method allows for quick recognition across a large database of stored faces or targets by computing only a subset of scores. This technique is particularly useful for applications with a large database or list, as well as applications that require processing large amounts of data, as in video surveillance applications. The face recognition and tracking algorithm, along with the fast search approach, can be used for video surveillance applications such as detecting, identifying, and tracking individuals and objects (such as suitcases or weapons), face recognition for physical and logical access control, watch list identification, suspicious behavior detection, and other applications, including providing real-time information about the people within a given operating environment. It could also be used to enhance SA in simulated AugCog environments by providing detection, identification, and tracking of humans in the visual environment.

4 A New Approach for Biometric Monitoring

The framework for this approach is similar to the edit distances used in text searches (and other searches) where the data can be represented by a one dimensional string of symbols, such as letters and numbers. Unlike traditional image recognition algorithms, the features extracted from the data are not used for classification. Instead, a mapping is found between the image to be identified and the database of faces, and properties of this mapping are used for classification. The mappings represent the edit distances previously only associated with one-dimensional data such as text [8, 9] and other data such as DNA [10] that is typically represented by a string of symbols. Once the mappings are found between the test and train images, properties associated with edit distances, namely – insertion, deletion, and substitution errors [8, 9] – are used to measure the degree of similarity between the images.

This approach involves the innovative mapping of two-dimensional image data into insertion, deletion, and substitution errors traditionally associated with one-dimensional strings [11,12 and 13]. Also, techniques used to solve the problem in one dimension, such as dynamic programming, cannot be extended to two-dimensional problems. In this unique approach, properties of the two-dimensional mapping between two images are found which represent deletions, insertions, and substitution (or match) distances in the one-dimensional problem. This technique is robust to variations in lighting and poses, which is critical for applications where surveillance cameras are used and the capture of video and still image data occurs in an uncontrolled environment. An early version of this algorithm was applied to the problem of face recognition [14] as well as image preprocessing and registration for face recognition [15].

The block matching algorithm used for motion estimation in current video coding standards such as MPEG [16, 17, 18, and 19] is the basic framework for the mapping between the database or train image and the unknown or test image. The mapping between test and train images can be applied in a "forward" direction where the mapping is found which converts the test image into the train image, as well as in a "backward" direction where the mapping is found that converts the train image into the test image. The forward and backward mappings are not simply the inverse of one another. The block matching algorithm was first introduced to perform motion estimation and compensation for video compression in order to take advantage of temporal correlations between video frames by estimating the current frame from the previous frame. In a hybrid video coder based on the traditional motion compensation scheme, motion estimation is performed by matching blocks [16] between the original frame and the previously reconstructed frame. An estimate of the current block can be obtained by searching similar blocks in the previous encoded (or original image) frame in a predetermined search area. The block matching algorithm is used for motion estimation between two video frames for compression and, in our case, the block matching algorithm is used for disparity estimation between the test image and each train image. In addressing face recognition, the key differentiator is that we expect the disparity map between the correctly matched faces to have significantly different properties than the disparity maps found for mismatched faces. We use the properties of the disparity fields found in mapping the test image and train images and how they relate to the traditional edit distances used in text for optical character recognition (OCR).

5 Applications of the P-Edit Distance to Face Recognition

While current applications of this face recognition technology include physical and logical access control and face detection, identification, and tracking for surveillance applications, STAR Face system has the potential to enhance the in SA in simulated AugCog environments by providing detection, identification, and tracking of the human element in the visual environment. Figures 3 and 4 illustrate typical surveillance video footage and face detection results for the STAR Face Recognition System.

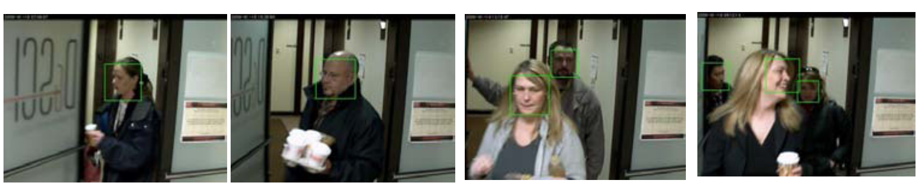

Fig. 3. Face detection results for surveillance data

Star Technologies Surveillance System detects faces from surveillance video to enroll individuals into a database and to recognize them in later footage. When an individual is detected for the first time and a match is not found in the current stored database, the database is updated with a new entry for that individual. Later occurrences of the same individual, either from the same sensor or other sensors, can then be identified. Other information about the individual can also be stored in the database for future reference.

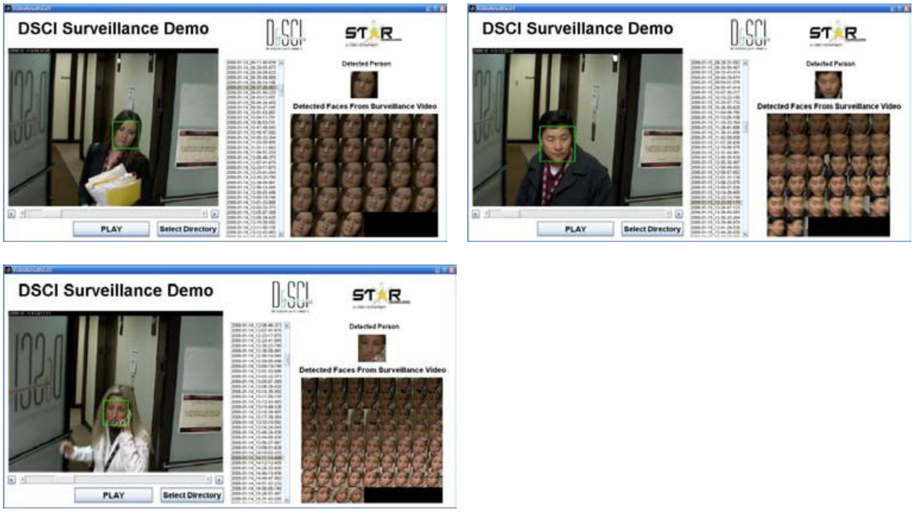

Fig. 4. User interface for the surveillance system and face detection results

6 Summary

Since human cognition is influenced by interaction with the physical environment, in order to measure user state for AugCog intervention, tools to enhance environmental monitoring abilities are needed. Emerging video surveillance and facial recognition/analysis technology has the potential to provide detection, identification, and tracking of targets in the operator's live environment to increase the accuracy of the AugCog system's automated SA. A new methodology for image-based face or object recognition based on extending the concept of edit distances for one-dimensional signals can be extended to the detection, identification, and tracking of individuals in a video sequence and includes a method for detecting and recovering from occlusions. A fast search method allows for fast recognition when the database of individuals to be identified is large or if the amount of data to be processed is large such as in video surveillance applications. The searches can be performed on still images, videos, or a combination of both. Boolean operators are supported so that the user can narrow down the search and filter out unwanted results. Facial recognition technology can be a critical solution for providing the necessary environmental SA for effective diagnosis and driving of AugCog mitigation, particularly in live operating environments.

References

1. Kruse, Schmorrow: Foundations of Augmented Cognition, 1st edn., pp. 441–445. Lawrence Erlbaum Associates, Mahwah (2005)
2. Schmorrow, Reeves: Foundations of Augmented Cognition, 2nd edn., p. XIII. Springer, Germany (2007)

3. Dorneich, et al.: Supporting real-time cognitive state classification on a mobile individual. JCEDM 1(3), 241–242 (2007)
4. Nicholson, et al.: An adaptive system for improving and augmenting human performance. In: Schmorrow (ed.) Foundations of Augmented Cognition, 2nd edn., pp. 215–222. Lawrence Erlbaum Associates, Mahwah (2006)
5. van Maanen, et al.: Closed-loop adaptive decision support based on automated trust assessment. In: Foundations of Augmented Cognition, 2nd edn., p. 267. Lawrence Erlbaum Associates, Mahwah (2007)
6. Shanker, T., Richtel, M.: In: new military, data overload can be deadly. New York Times (January 16, 2011)
7. Parasuraman, Rizzo: Introduction to neuroergonomics. In: Neuroergonomics: The Brain at Work, p. 6. Oxford University Press, New York (2007)
8. Wagner, R.A., Fisher, M.J.: The string-to-string correction problem. Journal of ACM 21, 168–178 (1974)
9. Gusfield, D.: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press, New York (1997)
10. Altschul, S.: Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219, 555–565 (1991)
11. Podilchuk, C., Barinov, L., Hulbert, W., Jairaj, A.: Face Recognition in a Tactical Environment. In: IEEE Proceedings of MILCOM (2010)
12. Podilchuk, C., Hulbert, W., Flachsbart, R., Barinov, L.: Face Recognition for Uncontrolled Environments. In: Proceedings of SPIE Defense, Security and Sensing, Biometric Technology for Human Identification VII, Orlando, Florida, April 5-9 (2010)
13. Hulbert, W., Podilchuk, C., Mammone, R.J.: Face Recognition Using a Pictorial-Edit Distance. In: Proceedings of the IEEE International Conference on Image Processing, ICIP 2008, pp. 1908–1911 (2008)
14. Podilchuk, C., Patel, A., Harthattu, A., Anand, S., Mammone, R.: A New Face Recognition Algorithm using Bijective Mappings. In: Proceedings IEEE International Conference on Computer Vision and Pattern Recognition (PAMI 2005), June 20-26, vol. 3, pp. 165–175 (2005)
15. Savvides, M., Xie, C., Chu, N., Kumar, B.V.K.V., Podilchuk, C., Patel, A., Harthattu, A., Mammone, R.: Robust face recognition using advanced correlation filters with bijective-mapping preprocessing. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 607–616. Springer, Heidelberg (2005)
16. Sikora, T.: MPEG digital video-coding standards. IEEE Signal Processing Magazine 14(5), 82–100
17. Servetto, S.D., Podilchuk, C.I.: Stochastic modeling and entropy constrained estimation of motion from image sequences. In: Proceedings of the IEEE International Conference on Image Processing (ICIP 1998), vol. 3, pp. 591–595 (1998)
18. Han, S.-C., Podilchuk, C.I.: Efficient encoding of dense motion fields for motion-compensated video compression. In: Proceedings of the IEEE International Conference on Image Processing (ICIP 2000), vol. 1, pp. 84–88 (2000)
19. Han, S.-C., Podilchuk, C.I.: Modeling and coding of DFD using dense motion fields in video compression. IEEE Transactions on Image Processing 10(11), 1605–1612 (2001)