



PERGAMON

Available online at www.sciencedirect.com

 ScienceDirect

Acta Astronautica 60 (2007) 341–350

ACTA
ASTRONAUTICA

www.elsevier.com/locate/actaastro

Monitoring of facial stress during space flight: Optical computer recognition combining discriminative and generative methods

David F. Dinges^{a,*}, Sundara Venkataraman^b, Eleanor L. McGlinchey^a,
Dimitris N. Metaxas^b

^aUnit for Experimental Psychiatry, Department of Psychiatry, University of Pennsylvania School of Medicine, 423 Guardian Drive, 1013 Blockley Hall, Philadelphia, PA 19104-6021, USA

^bDepartment of Computer Science, Center for Computational Biomedicine, Imaging and Modeling, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA

Available online 17 October 2006

Abstract

Astronauts are required to perform mission-critical tasks at a high level of functional capability throughout spaceflight. Stressors can compromise their ability to do so, making early objective detection of neurobehavioral problems in spaceflight a priority. Computer optical approaches offer a completely unobtrusive way to detect distress during critical operations in space flight. A methodology was developed and a study completed to determine whether optical computer recognition algorithms could be used to discriminate facial expressions during stress induced by performance demands. Stress recognition from a facial image sequence is a subject that has not received much attention although it is an important problem for many applications beyond space flight (security, human–computer interaction, etc.). This paper proposes a comprehensive method to detect stress from facial image sequences by using a model-based tracker. The image sequences were captured as subjects underwent a battery of psychological tests under high- and low-stress conditions. A cue integration-based tracking system accurately captured the rigid and non-rigid parameters of different parts of the face (eyebrows, lips). The labeled sequences were used to train the recognition system, which consisted of generative (hidden Markov model) and discriminative (support vector machine) parts that yield results superior to using either approach individually. The current optical algorithm methods performed at a 68% accuracy rate in an experimental study of 60 healthy adults undergoing periods of high-stress versus low-stress performance demands. Accuracy and practical feasibility of the technique is being improved further with automatic multi-resolution selection for the discretization of the mask, and automated face detection and mask initialization algorithms.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Optical computer recognition; Computer vision; Stress; Human face; Astronauts; Markov models

1. Introduction

The measurement of facial expressions is a well studied problem in the fields of anthropology, psychology

[1,2] and more recently computer vision [3–9]. Facial expressions are an important factor in understanding the emotional state of humans and even more so in the case of humans performing critical tasks while under stress. Astronauts are required to perform mission-critical tasks at a high level of functional capability throughout spaceflight. While they can be trained to cope with, and/or adapt to some stressors of spaceflight, stressful reactions

* Corresponding author. Tel.: +1 215 898 9949;
fax: +1 215 573 6410.

E-mail address: dinges@mail.med.upenn.edu (D.F. Dinges).

can and have occurred during long-duration missions, especially when operational performance demands become elevated when unexpected and/or underestimated operational requirements occur while crews are already experiencing work-related stressors [10]. In some of these instances, stressed flight crews have withdrawn from voice communications with ground controllers or, when pressed to continue performing, they made errors that could have jeopardized the mission. Consequently, there is a need to identify when during operational demands astronauts are experiencing behavioral stress associated with performance demands. This is especially important as mission durations increase in length and ultimately involve spaceflight to other locations in the solar system.

Recognizing stress in human faces in spaceflight also may prove to be a potentially valuable tool for safety-sensitive activities on Earth, such as security, human-machine interface, and fitness for duty evaluations. The pioneering studies of Ekman and colleagues [1,2] helped establish that specific human emotions are expressed similarly in faces of different ethnicities. The advent of computer vision methods has led to devising methods to track facial features and automatically measure their movement [3–9]. These measurements typically seek to detect emotions, but not their relationship to performance. An alternative to attempting to use computer vision to identify emotional states is to develop it to recognize when someone is showing signs of escalating stress during critical task performance. This approach allows optical computer recognition to be developed idiosyncratically such that it can learn to discriminate stress in facial expressions that may be unique to a given astronaut based on ethnicity, culture, gender, or personality. Computer vision of facial expressions, therefore, offers a completely obtrusive way to monitor behavioral stress that may compromise performance during the spaceflight. In this article we detail for the first time, the mathematical development of a novel optical computer recognition system for detecting behavioral stress, and update its current experimental validity [10].

2. Methods

To evaluate the validity of the optical computer recognition approach we developed, an experiment was conducted that varied stress reactions using performance demands in $N = 60$ healthy adults (mean age = 30 ± 6.8 yr). Body mass index for all subjects was within 15% of normal. In order to participate, individuals needed to have a stable, normally timed sleep-wake cycle; be

free of alcohol or drug abuse; and be non-smokers. Subjects with current depression as determined by the Beck Depression Inventory were screened out of the study. Eyewear was not allowed (except for contact lenses), as glasses obscured facial features that must be visualized by the optical recognition system. The experimental procedures and informed consent were approved by the Institutional Review Board of the University of Pennsylvania.

2.1. Procedures

A preliminary published report [10] of the optical computer recognition algorithm accuracy contains full details of the experimental procedures. During the in-laboratory session, subjects completed questionnaires measuring aspects of personality, mood, and stress perception and coping strategies, to determine if variations in these psychological parameters related to the identification of facial expressions of stress. These scales included the Toronto Alexithymia Scale [11]; State/Trait Anxiety Index [12]; Depression Inventory; Perceived Stress Scale [13]; Life Orientation Test [14]; The COPE Inventory [15]; Distress Inventory; Morningness/Eveningness Questionnaire [16]; Millon Index of Personality Styles [17]; Rotter's Locus of Control [18]; Pittsburgh Sleep Quality Index [19]; Marlowe-Crowne Social Desirability Scale [20]; and Penn State Worry Questionnaire [21].

Upon arriving at the laboratory, subjects had EKG electrodes attached; saliva collection procedures explained; practiced the performance tasks; and completed several questionnaires. We used standardized performance tasks presented from a computerized neurobehavioral test battery that we developed and validated in previous experiments [22,23]. Videos of the face were acquired during cognitive performance testing, and stress reactions were tracked during both low- and high-workload stressor conditions using self-report visual analog ratings and mood scores.

Subjects were studied twice during both low- and high-workload stressor performance demands—once to provide data to train the optical computer recognition algorithm on typical facial expressions of each subject during low- and high-workload demands, and once to provide data to “test” the algorithm's accuracy in predicting which facial video was low workload and which was high workload. High-workload scenarios induced greater stress by involving more difficult performance tasks, negative social feedback, and greater time pressure relative to low-workload scenarios. The following tasks were administered to subjects in easy versions

during the low-workload test bouts and in difficult versions during the high-workload test bouts: the Stroop word–color interference task [24], the psychomotor vigilance task [25], the probed recall memory [26], the descending subtraction task [27], the visual memory task, the digit symbol substitution task, the serial addition subtraction task [28], the synthetic workload task [29], the meter reading task, the logical reasoning task [30], and the Hayling sentence completion task [31]. High-workload versions of the tasks required more attention, skill and faster responding.

In addition, stress was further reduced or induced by false computer feedback, respectively, for the low- and high-workload conditions. That is, periodically throughout low- and high-workload test bouts, computer feedback was provided to subjects. During the low-workload periods, this feedback was very positive regarding performance level relative to other subjects (i.e., low-stressor scenario). During the high-workload periods computer feedback was very negative regarding performance level relative to other subjects (i.e., high-stressor scenario). The nature of the social and computer-generated feedback within each type of stressor scenario was identical for all subjects and independent of subjects' actual performance levels.

Stress reactions were tracked using self-report ratings of stress and mood states. Before and after every workload bout, subjects responded to visual analog scales for alertness, mental and physical fatigue, exhaustion and stress. Following each workload bout, subjects completed the Profile of Mood States [32], an adjective checklist for assessing six dimensions of mood (fatigue–inertia, vigor–activity, confusion–bewilderment, tension–anxiety, anger–hostility, depression–dejection) and overall mood disturbance. Following every task, subjects also completed visual analog scales for rating distress, effort, frustration and difficulty of tasks along with pre- and post-test stress questionnaires assessing subjects' perception of how stressed they feel at this point in time. Results confirming the validity of the low- versus high-stress nature of these performance tasks are presented elsewhere [10].

2.2. Development of the optical computer recognition algorithm and deformable models

The study design consisted of an algorithm training phase and an algorithm testing/recognition phase. During the training phase, the training sequences were tracked using the tracker. The resulting parameters were used to train the hidden Markov models (HMMs) after suitable modification. During the

testing/recognition phase, the parameters from tracking the testing sequences were used to test for high/low stress. The following sections provide an overview of tracking and testing/recognition methods relative to optical computer recognition algorithm development.

2.2.1. Face tracking using deformable models

This section outlines the theory of deformable models and the face tracking technique used. The face model was initialized by first fitting it to the person's face in the first frame of the sequence. The video was then tracked statistically by cue integration [33]. The tracking produced a vector of parameters which defined the translation, rotation and deformation parameters for each frame.

Face tracking was accomplished by statistically integrating visual cues from the images to get forces acting on our deformable face model. The model was deformed by a parameter vector \mathbf{q} , which is responsible for rigid and non-rigid motions [34]. The non-rigid motions consisted of deformations that characterize a face, such as raising eyebrows, lip stretching, lip curving and jaw movements. An image point p_i that is on the model was mapped to the parameter space as follows:

$$p_i = F_i(\mathbf{q}). \quad (1)$$

Tracking a frame then involved finding the optimal parameter vector \mathbf{q} that achieved the correspondence between the model and the image. In the general case, the mapping F_i could be non-linear and a closed-form solution might not exist. Computer vision algorithms were used to get approximate correspondences. These algorithms yielded 2D displacements \mathbf{f}_i that we designated as *image forces*, to adjust the parameters \mathbf{q} from the previous frame to the current frame. The parameters \mathbf{q} were obtained by an optimization procedure that minimized the sum of the magnitude of all the image forces \mathbf{f}_i .

The image forces \mathbf{f}_i were then mapped into a *generalized force* where B_i is the projected,

$$\mathbf{f}_g = \sum_i B_i^T \mathbf{f}_i, \quad (2)$$

Jacobian of the model at the point p_i to which the image force \mathbf{f}_i is applied. Using the generalized force \mathbf{f}_g we solved the dynamical system where K is the stiffness matrix, using a

$$\mathbf{q} = K \mathbf{q} + \mathbf{f}_g \quad (3)$$

simple gradient descent method. We assumed that we had one single generalized force for all the cues, but in practice the cues were dissimilar. For example, the point

tracker worked best on regions with complex texture, while the edge tracker worked best on high contrast regions. In order to effectively integrate all the cues, a statistical framework was needed. For each cue the 2D image force \mathbf{f}_{cj} was represented as affine forms as follows:

$$\hat{\mathbf{f}}_{cj} = \mathbf{f}_{cj,0} + \sum_{i=1}^m \mathbf{f}_{cj,i} \varepsilon_i, \quad (4)$$

where the coefficients $\mathbf{f}_{cj,i}$ are 2D vectors and ε_i are the symbolic real variables whose values are known to lie in the interval $[-1, 1]$. The central quantity $\mathbf{f}_{cj,0}$ was designated the *central value*, which is analogous to the mean of an affine distribution, and ε_i were considered the *noise variables*. Each noise variable ε_i represents an independent component of the overall uncertainty. Here, \mathbf{f}_{cj} represents the region of uncertainty around the image force as a convex polygon, whose number of faces depend on m .

The cue's generalized force was the sum of all the contributions, after we projected them onto the parameter space, using the Jacobian of the deformable model at each point:

$$\hat{\mathbf{f}}_{gc} = \sum_j \mathbf{B}_j^T \hat{\mathbf{f}}_{cj}, \quad (5)$$

where \mathbf{B}_j^T is the projected Jacobian of the model at point j and $\hat{\mathbf{f}}_{gc}$ is the image force that the cue c applies to the point j . Thus, the above equation resulted in an n -dimensional affine form that represented the cue's generalized form. We could make an independence assumption about all the 2D forces if they were reasonably far apart, leading to the independence of the noise variables $\hat{\mathbf{f}}_{gc}$. Using *Lindeberg's theorem* it was possible to prove that the sum would converge to a Gaussian distribution $\tilde{\mathbf{f}}_{gc}$ with mean μ_c and covariance matrix A_c . We estimated the mean and covariance directly by applying expectations to Eq. (5) above. The mean vector was

$$\mu_c = E[\hat{\mathbf{f}}_{gc}] = \mathbf{f}_{g,c,0}$$

and the covariance matrix was

$$A_c = E[(\hat{\mathbf{f}}_{g,c} - \mathbf{f}_{g,c,0})(\hat{\mathbf{f}}_{g,c} - \mathbf{f}_{g,c,0})^T]$$

which resulted in

$$\lambda_i j = \sum_{k=1}^m \mathbf{f}_{gck_i} \mathbf{f}_{gck_i} E[\varepsilon^2 k]. \quad (6)$$

We then used *maximum likelihood estimation* to find the optimal \mathbf{f}_g to identify the final generalized force.

The final generalized force drives the deformations on the model by combining the generalized forces from each of the cues [33].

2.2.2. Recognition

Hidden Markov models are logical choices for modeling temporal processes. However, they require considerable amounts of training data. Since there had been no previous work in visual stress recognition, it was unclear how big the training dataset had to be and what facial features needed to be used for recognition. While we generated a sizable experimental dataset from our experiment for this training, a more reliable technique in addition to HMMs would be better for classification. Consequently, we used support vector machines (SVMs), which refer to a static classification technique that is highly popular. SVMs have the advantage of being non-linear and geared specifically towards classification. This is achieved by casting the training as reducing the margin of classification error. Another motivation for using a static classification technique on top of HMMs was to appropriately segment the output probabilities from the HMMs. The output probabilities from the HMMs in our case were sometimes very close for high- and low-stress data. In such cases we could proceed with the label that had a higher probability. Nevertheless, a better discriminative technique would be more useful and would give a clear decision boundary between the high- and low-stress probabilities.

2.2.3. Hidden Markov models

HMMs are an ideal tool to model temporal processes. An HMM λ consists of N states S_1, S_2, \dots, S_N . It includes transitions from one state to another, and switches from the current state to a new state at regular time intervals. The probability of going from state S_i to S_j is defined by a_{ij} . The probability of the system starting at state S_i is defined as p_i . If the number of observation symbols are M , then we can define $b_j(k)$ as the probability that a given symbol k is generated at time instance j .

Now, given the above data we have the following three problems [35,36]:

- (1) Given the observation sequence O_1, O_2, \dots, O_T , and a model λ , how do we efficiently compute $P(O|\lambda)$? In other words, what is the probability that the model λ generated this sequence of observations?
- (2) Given the observation sequence O and the model λ , what is the state sequence Q that maximizes $P(Q, O|\lambda)$? That is, given the model and the observations, what is the sequence of states that generated these observations?

(3) Given an HMM λ , how should the transition, observation and initial probabilities be adjusted?

The first problem corresponded to recognizing an unknown data sequence with the high-stress and low-stress HMMs. The probability that each HMM generated the unknown sequence $P(O|\lambda)$ was computed. The video of a subject's face was then classified as being either high or low stress, depending on which HMM had a higher probability. The computation of $P(O|\lambda)$ was as follows:

Let $Q = Q_1, Q_2, \dots, Q_T$ be a state sequence in the HMM λ . Let

$$\alpha_i(i) = P(O_1, O_2, \dots, O_t, Q_t = S_i | \lambda), \quad 1 \leq i \leq N, \quad (7)$$

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i), \quad (8)$$

$$\alpha_1(i) = \pi_i b_i(O_1), \quad (9)$$

$$\alpha_{t+1}(i) = b_i \left(O_{t+1} \sum_{j=1}^N \alpha_t(j) a_{ji} \right), \quad 1 \leq t \leq T - 1. \quad (10)$$

It should be noted that these equations assume that the O_i are independent, and they make the Markovian assumption that the transition to the next state is only dependent on the current state. This method is called the “forward–backward” procedure and computed $P(O|\lambda)$ in $O(N^2T)$ time.

The second problem corresponded to finding the most likely path through an HMM λ , given an observation sequence O , and is equivalent to maximizing $P(Q, O|\lambda)$. Let

$$\delta_t(i) = \max_{Q_1, \dots, Q_{t-1}} P(Q_1, Q_2, \dots, Q_t = S_i, O|\lambda), \quad (11)$$

$$\delta_{t+1}(i) = b_i(O_{t+1}) \max_{1 \leq j \leq N} \{\delta_t(j) a_{ji}\}, \quad (12)$$

$$P(Q, O|\lambda) = \max_{1 \leq i \leq N} \{\delta_T(i)\}. \quad (13)$$

The $\delta_t(i)$ corresponds to the maximum probability of all state sequences that end up in state S_i at time t . The Viterbi algorithm is a dynamic programming algorithm that, using the above equation, computes both the maximum probability $P(Q, O|\lambda)$ and the state sequence Q in $O(N^2T)$ time.

The third problem corresponded to training the HMMs with data, such that they were able to recognize previously unseen data correctly after the training

phase. There exists no analytical solution for maximizing $P(O|\lambda)$ for given observation sequences, but an iterative procedure called the Baum–Welch procedure maximizes $P(O|\lambda)$ locally. In the case of continuous density output probabilities the procedure works as follows:

Define $b_j(O)$ as $b_j(O) = \sum_{m=1}^M c_{jm} G(O, \mu_{jm}, U_{jm})$, where M describes the number of mixtures, j is the state number, c describes the weight of the mixture m in state j , G is a Gaussian density with mean μ and covariance matrix U . Define the backward variable β as

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | Q_t = S_i, \lambda), \quad (14)$$

$$\beta_T(i) = 1, \quad (15)$$

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1} \beta_{t+1}(j)), \quad (16)$$

$$1 \leq i \leq N, \quad 1 \leq t \leq T - 1. \quad (17)$$

Furthermore, define ξ and γ as

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)}, \quad (18)$$

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j). \quad (19)$$

The $\sum_t \xi_t(i, j)$ can be interpreted as the expected number of transitions taken from S_i . With these interpretations, the re-estimation formulae for the transitions and output probabilities were

$$\bar{\pi}_i = \gamma_1(i), \quad (20)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad (21)$$

$$\bar{c}_{jm} = \frac{\sum_{t=1}^T \gamma_t(j, m)}{\sum_{t=1}^T \sum_{k=1}^M \gamma_t(j, k)}, \quad (22)$$

$$\bar{\mu}_{jm} = \frac{\sum_{t=1}^T \gamma_t(j, m) O_t}{\sum_{t=1}^T \gamma_t(j, m)}, \quad (23)$$

$$\bar{U}_{jm} = \frac{\sum_{t=1}^T \gamma_t(j, m) (O_t - \mu_{jm})(O_t - \mu_{jm})^T}{\sum_{t=1}^T \gamma_t(j, m)}. \quad (24)$$

Repeated use of this re-estimation procedure converged to the maximum probability. An important thing

to note here is the number of training samples that might be needed to effectively model the process. The number of training samples needs to be quite large and is the main drawback for HMM-based recognition techniques.

2.2.4. Support vector machines

SVMs are a very efficient means for classifying data that is not linearly separable [37,38]. The key observation is that the decision boundary or “margin” should be as far away from the data as possible. We let $x = \{x_1, x_2, \dots, x_n\}$ be our dataset and $y_i \in \{1, -1\}$ be the class label of x_i . If the margin is w , then in order to classify all the data points correctly, $y_i (w^T x_i + b) > 1$. Then, we have a constrained minimization problem $\min \frac{1}{2} \|w\|^2$ subject to $y_i (w^T x + b) > 1$.

Now, consider the dual of this minimization

$$\begin{aligned} \max \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n i - 1, \\ & j - 1^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t.} \quad \alpha_i &\geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \tag{25}$$

This is a quadratic programming problem and the global maximum can always be found. Also, w can always be recovered by $w = \sum_{i=1}^n \alpha_i y_i x + i$.

Most of the α_i is zero and w is a linear combination of a small number of examples. The x_i with non-zero α_i are called support vectors. The decision boundary is determined only by the support vectors. If t_j , ($j = 1, \dots, s$) are the indices of the support vectors, $w = \sum_{j=1}^s \alpha_{t_j} y_{t_j} x_{t_j}$. For testing with new data z , compute

$$w^T z + b = \sum_{j=1}^s \alpha_{t_j} y_{t_j} x_{t_j}^T z + b \tag{26}$$

and classify as class 1 if it is positive, class 2 otherwise.

If classification error $\xi_i = 0$, then there is no error for x_i . Then we have the following conditions:

$$\begin{cases} w^T x_i + b \geq 1 - \xi_i, & y_i = 1, \\ w^T x_i + b \leq -1 + \xi_i, & y_i = -1, \\ \xi_i \geq 0 & \forall i. \end{cases}$$

Now the minimization problem can be recast as $\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$ subject to $y_i (w^T x + b) \geq 1 - \xi_i$, $\xi_i \geq 0$.

So the dual of the problem would be

$$\begin{aligned} \max \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n i - 1, \\ & j - 1^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t.} \quad C &\geq \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \tag{27}$$

This margin minimization provided an explicit measure of classification error. This idea was extended to non-linear decision boundaries by the “kernel trick”. The input data x_i was mapped to a higher dimension in such a way that a linear operation in the higher dimension would be equivalent to a non-linear operation in the input dimension. This seems counter-intuitive since the mapping operation might be computationally expensive and might be hard to estimate. Choosing the right kernel reduces computation and minimizing $\|w\|^2$ leads to a good estimate.

Consider the following kernel example:

$$K(x, y) = (1 + x_1 y_1 + x_2 y_2)^2. \tag{28}$$

Now consider the following transformation:

$$\phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1 x_2). \tag{29}$$

Now, the inner product between $\phi(x)$ and $\phi(y)$ will be:

$$\langle \phi(x), \phi(y) \rangle = (1 + x_1 y_1 + x_2 y_2)^2 = K(x, y). \tag{30}$$

So the inner product can be computed without going through the kernel function. Going back to (weights) and using the kernel functions we get

$$\begin{aligned} \max \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n i - 1, \\ & j - 1^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad C &\geq \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \tag{31}$$

Now, the weights become $w = \sum_{j=1}^s \alpha_{t_j} y_{t_j} \phi(x_{t_j})$.

During testing, from (test), we get

$$f = \langle w, \phi(z) \rangle + b = \sum_{j=1}^s \alpha_{t_j} y_{t_j} K(x_{t_j}, z) + b. \tag{32}$$

Thus we obtained non-linear decision boundaries between data points for classification. In our application, some of the facial video sequences could be easily classified as high or low stress since the output

probabilities from the HMMs were almost the same for both high- and low-stress situations. By having a non-linear decision boundary, the classification was somewhat more robust in such situations.

3. Results

Fig. 1 reveals that subjects' ratings of general stress level during performance tests increased significantly

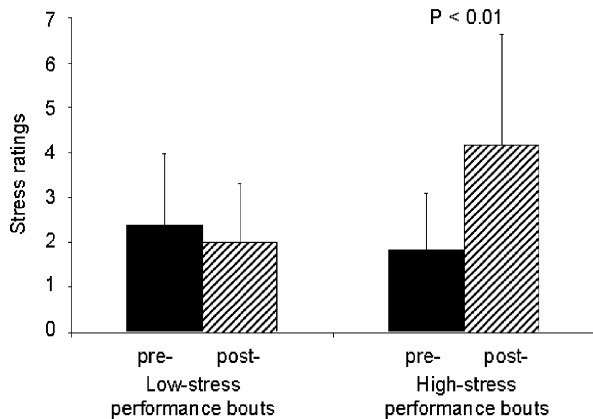


Fig. 1. Mean \pm sem of ratings of stress made by $N = 60$ healthy adults using a 10 cm visual analog scale prior to (solid bars) and immediately following (hatched bars) low-stress performance test bouts (i.e., easy tasks and positive feedback) and high-stress performance test bouts (i.e., difficult tasks and negative feedback). A higher score represents greater perceived stress. Only the high-stress performance test bouts evoked significant elevation of perceived stress ($P < 0.01$).

following completion of high-workload (stressor) test bouts compared to low-workload test bouts. Virtually identical profiles were found for ratings of difficulty of tasks, effort required to perform, and frustration [10]. In all cases, high-workload performance demands and their accompanying negative feedback significantly increased these negative experiences indicative of stress, relative to low-workload demands and their accompanying positive feedback. Fig. 2 shows an example of the current optical computer algorithm mask applied to the face of a subject in a low-stress testing session. Progressive improvements described above in programming of the deformable face mask increased its quality, properties, and precision. Fig. 3 displays an example of two facial expressions during a high-stress performance (testing) session. Next to each example are graphs illustrating the quantitative detection of facial expressions involving stress-related changes in the area of the mouth. These graphs show that the deformations from the optical computer recognition system track changes in measurement around the mouth over time.

To develop the current algorithm, images were normalized (i.e., the inverse of the rigid transformations inferred from the deformable model tracking were applied to the face model) to produce a frontal face with the image from the video sequence texture mapped onto this model. From the large number of video sequences that were analyzed from the 60 subjects, a number of common indicators of high stress were identified in specific areas of facial expression, especially the eyebrows

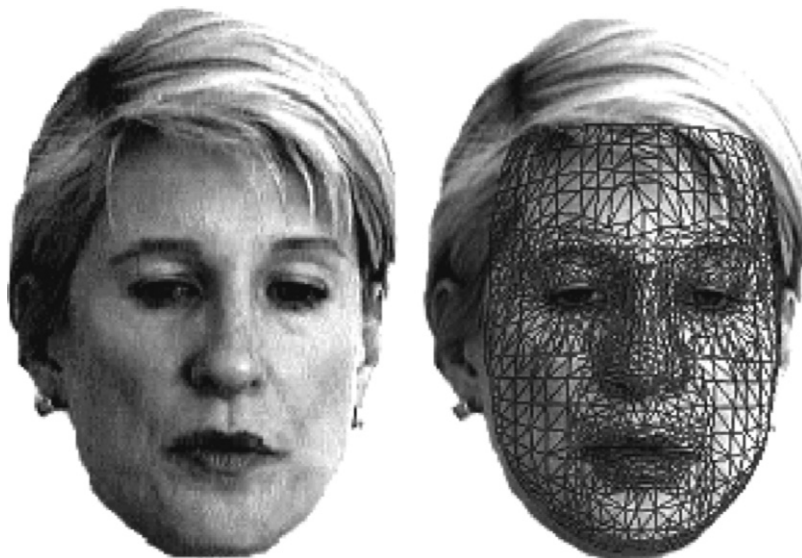


Fig. 2. Photo example of the current optical computer algorithm mask applied to the face of a subject in a low-stress testing session. Progressive improvements in the programming of the deformable face mask have increased its quality, properties, and precision. Subject provided written permission for display of facial image.

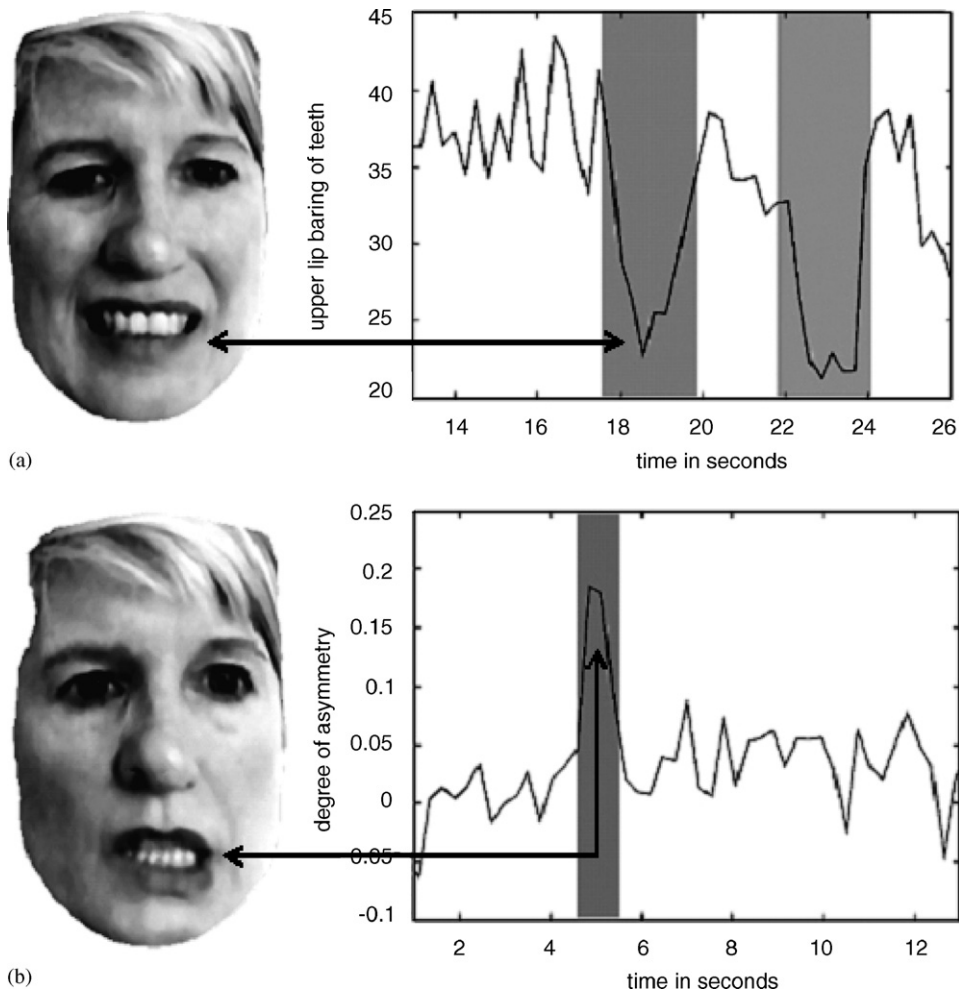


Fig. 3. A number of common indicators of high stress were identified in specific areas of facial expression, especially the eyebrows and mouth areas. Examples of the latter are shown here. Fig. 3a shows a subject baring teeth during the high-workload stressor. The graph on the right of Fig. 3a is a representation of lip movement displayed by the subject. It quantifies lip distance plots into the slopes of the best-fit lines to the curves over time. Strong negative spikes in the intercept denote an upward movement of the upper lip, toward the nose, and display of the teeth. The highlighted regions (gray bars) correspond to the stressed expression in photo 3a, as indicated by arrows. Fig. 3b shows an example of asymmetric lip deformation during the high-workload stressor in the same subject. The graph to the right of Fig. 3b quantifies the degree of asymmetry of the lower lip. The highlighted region (gray bar) corresponds to the stressed expression in photo 3b. Subject provided written permission for display of facial image.

and mouth areas. Fig. 3 shows two of these expressions from the mouth area during high stress: baring of teeth (Fig. 3a) and asymmetric lip deformations (Fig. 3b), both of which could be detected by the current algorithm using the statistical cue integration method. Such facial deformations formed the basis for training the current optical computer recognition system.

Video sequences for all 60 subjects were used in the algorithm training and recognition phase of the study. Four video sequences were used for each subject, two for training the algorithm for each subject (one from a low-stress condition and one from a high-stress

condition) and two for testing the algorithm's accuracy in discriminating low- from high-stress conditions. There were a total of 240 training video sequences and 60 testing video sequences. These sequences were annotated for various facial movements (typically eyebrow regions and mouth regions), and more recently hand movements to the face have been added, as the most promising ways to identify stress during performance test bouts. The training and testing video sequences were extracted from the original sequences using the annotations provided. These video sequences were typically around 5–10 s long. The training and

testing videos were then tracked using our statistical cue integration method. The parameters obtained from the tracking were used to calculate the asymmetric features for training the HMMs.

The HMMs consisted of five hidden states. We used a holistic approach for recognition, classifying an entire video sequence as being a high- or low-stress situation. We used the hidden Markov toolkit (HTK) implementation of HMMs for this task. For SVMs, we used the BSVM implementation. Preliminary results on the accuracy of the optical computer algorithm indicate that the current version reliably discriminated high-stress from low-stress test conditions in 42 of the 60 subjects (70%). In the approximately 7% of the sequences in which the optical computer recognition algorithm failed to accurately discriminate, poor lighting conditions degraded tracking accuracy. Tracking also failed to discriminate a subject who had a very high score on the alexithymia scale. People who have alexithymia are unable to express emotions satisfactorily although they feel the emotion. Excluding poor lighting conditions for all subjects in which they occurred (regardless of algorithm accuracy), the current algorithm was 68% accurate. This is better than chance, but below the 85% accuracy of a human judge blind to conditions [10]. We are currently in the process of designing new and additional methods to increase optical computer recognition algorithm accuracy.

4. Discussion

The results illustrate the utility of the optical computer recognition methodology to identify performance-induced stress in human faces. We are currently investigating more novel methods of combining generative and discriminative methods to improve recognition accuracy to levels at or above a human rater. The next phase of the project will include more video data with use of stereo cameras. At present we are only using one video sequence per stress condition per person. In the future we will be using all the annotated data from the subjects for training and testing. Based on our results, we feel that the database we have created and the recognition system will provide researchers with an important tool to further research in building real-world systems for unobtrusive, automated detection of stress from video of human expressions.

Acknowledgments

This research was supported by NASA cooperative agreement NCC 5-98 with the National Space Biomedical Research Institute. We thank Siome K. Goldenstein

and Christian Vogler for contributing to optical computer recognition techniques and algorithm development; Jillian Dorrian, Robert L. Rider, Naomi L. Rogers and Ziga Cizman for participating in data acquisition; and Adrian Ecker, John W. Powell, IV, Michele Carlin, Christos Ballas and Olivera Crenshaw for making the project technically feasible.

References

- [1] P. Ekman, Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique, *Psychological Bulletin* 115 (2) (1994) 268–287.
- [2] P. Ekman, W.V. Friesen, *Facial Action Coding System: Investigator's Guide*, Consulting Psychologists Press, Palo Alto, CA, 1978.
- [3] I.A. Essa, A.P. Pentland, Coding, analysis, interpretation and recognition of facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (7) (1997) 757–763.
- [4] I. Cohen, A. Garg, T. Huang, Emotion recognition using Multi-level HMM, *NIPS Workshop on Affective Computing*, 2000.
- [5] I. Cohen, N. Sebe, A. Garg, T. Huang, Facial expression recognition from video sequences, *International Conference on Multimedia and Expo*, 2002.
- [6] I. Cohen, N. Sebe, A. Garg, T. Huang, Emotion recognition using a Cauchy Naive Bayes classifier, *International Conference on Pattern Recognition*, 2002.
- [7] I. Cohen, N. Sebe, L. Chen, A. Garg, T. Huang, Facial expression recognition from video sequences: temporal and static modelling, *Computer Vision and Image Understanding* 91 (1–2) (2003) 160–187.
- [8] T. Kanade, J. Cohn, Y. Tian, *Comprehensive database for facial expression analysis*, *IEEE Conference on Automatic Face and Gesture Recognition*, 2000.
- [9] Y. Tian, T. Kanade, J.F. Cohn, Recognizing action units for facial expression analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2) (2001) 97–115.
- [10] D.F. Dinges, R.L. Rider, J. Dorrian, E.L. McGlinchey, N.L. Rogers, Z. Cizman, S.K. Goldenstein, C. Vogler, S. Venkataraman, D.N. Metaxas, Optical computer recognition of facial expressions associated with stress induced by performance demands, *Aviation, Space and Environmental Medicine* 76 (6) Section II (2005) B172–B182.
- [11] G.J. Taylor, D. Ryan, R.M. Bagby, Toward the development of a new self-report alexithymia scale, *Psychotherapy and Psychosomatics* 44 (1985) 191–199.
- [12] C.D. Spielberg, R.L. Gorsuch, R.E. Lushene, *Manual for the State-Trait Anxiety Inventory*, Consulting Psychologists Press, Palo Alto, CA, 1970.
- [13] S. Cohen, G.M. Williamson, *Perceived Stress in a Probability Sample of the United States*, Sage, Beverly Hills, CA, 1988.
- [14] M.F. Scheier, C.S. Carver, Optimism, coping, and health: assessment and implications of generalized outcome expectancies, *Health Psychology* 4 (1985) 219–247.
- [15] C.S. Carver, M.F. Scheier, J.K. Weintraub, Assessing coping strategies: a theoretically based approach, *Journal of Personality and Social Psychology* 56 (1989) 267–283.
- [16] J.A. Horne, O. Ostberg, A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms, *International Journal of Chronobiology* 4 (1976) 97–110.

- [17] T. Millon, A theoretical derivation of pathological personalities, in: T. Millon, G. Klerman (Eds.), *Contemporary Directions in Psychopathology*, Guilford, New York, NY, 1986.
- [18] J.B. Rotter, Generalized expectancies for internal versus external control of reinforcement, *Psychological Monographs* 80 (1966) 609.
- [19] D.J. Buysse, C.F. Reynolds, T.H. Monk 3rd, S.R. Berman, D.J. Kupfer, The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research, *Psychiatry Research* 28 (1989) 193–213.
- [20] D.P. Crowne, D. Marlowe, A new scale of social desirability independent of psychopathology, *Journal of Consulting Psychology* 24 (1960) 349–354.
- [21] T.J. Meyer, M.L. Miller, R.L. Metzger, T.D. Borkovec, Development and validation of the Penn State Worry Questionnaire, *Behaviour Research and Therapy* 28 (1990) 487–495.
- [22] D.F. Dinges, Probing the limits of functional capability: the effects of sleep loss on short-duration tasks, in: R.J. Broughton, R.D. Ogilvie (Eds.), *Sleep, Arousal, and Performance*, Birkhäuser, Boston, 1992, pp. 177–188.
- [23] D.F. Dinges, N.B. Kribbs, Performing while sleepy: effects of experimentally induced sleepiness, in: T.H. Monk (Ed.), *Sleep, Sleepiness and Performance*, Wiley, Chichester, 1991, pp. 97–128.
- [24] J.R. Stroop, Studies of interference in serial verbal reactions, *Journal of Experimental Psychology* 18 (1935) 643–662.
- [25] D.F. Dinges, J.W. Powell, Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations, *Behavior Research Methods, Instruments, and Computers* 17 (1985) 652–655.
- [26] D.F. Dinges, N.B. Kribbs, B.L. Bates, M.M. Carlin, A very brief probed-recall memory task: sensitivity to sleep loss, *Sleep Research* 22 (1993) 330.
- [27] D.F. Dinges, E.C. Orne, F.J. Evans, M.T. Orne, Performance after naps in sleep conducive and alerting environments, in: L.C. Johnson, D.I. Tepas, W.P. Colquhoun, M.J. Colligan (Eds.), *Biological Rhythms, Sleep and Shift Work. Advances in Sleep Research*, vol. 7, Spectrum Publications, New York, NY, 1981, pp. 539–552.
- [28] D.R. Thorne, S.G. Genser, G.C. Sing, F.W. Hegge, The Walter Reed performance assessment battery, *Neurobehavioral Toxicology and Teratology* 7 (1985) 415–418.
- [29] T.F. Elsmore, Synwork1: a PC-based tool for assessment of performance in a simulated work environment, *Behavior Research Methods, Instruments, and Computers* 26 (1994) 421–426.
- [30] A. Baddeley, A 3-min reasoning task based on grammatical transformation, *Psychonomic Science* 10 (1968) 341–342.
- [31] P.W. Burgess, T. Shallice, Response suppression, initiation and strategy use following frontal lobe lesions, *Neuropsychologia* 34 (1996) 263–273.
- [32] D.M. McNair, M. Lorr, L.F. Droppleman, EITS Manual for the Profile of Mood States, Educational and Industrial Testing Service, San Diego, 1971.
- [33] S. Goldenstein, C. Vogler, D. Metaxas, Statistical cue integration in DAG Deformable Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (7) (2003) 801–813.
- [34] D. Metaxas, *Physics-based Deformable Models: Applications to Computer Vision, Graphics and Medical Imaging*, Kluwer Academic Publishers, Dordrecht, 1996.
- [35] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* 77 (2) (1989) 257–286.
- [36] C. Vogler, D. Metaxas, ASL recognition based on a coupling between hidden Markov models and 3D motion analysis, *International Conference on Computer Vision*, 1998.
- [37] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [38] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* 2 (2) (1998).