Optical computer recognition of facial expressions associated with stress induced by performance demands

Article *in* Aviation Space and Environmental Medicine · July 2005

Source: PubMed

10 authors, including:

David Dinges
University of Pennsylvania
367 PUBLICATIONS 33,133 CITATIONS

Jillian Dorrian
University of South Australia
178 PUBLICATIONS 3,225 CITATIONS

Eleanor McGlinchey
New York State Psychiatric Institute
36 PUBLICATIONS 912 CITATIONS

Naomi L Rogers
The University of Sydney
85 PUBLICATIONS 3,471 CITATIONS

Optical Computer Recognition of Facial Expressions Associated with Stress Induced by Performance Demands

David F. Dinges, Robert L. Rider, Jillian Dorrian, Eleanor L. McGlinchey, Naomi L. Rogers, Ziga Cizman, Siome K. Goldenstein, Christian Vogler, Sundara Venkataraman, and Dimitris N. Metaxas

Dinges DF, Rider RL, Dorrian J, McGlinchey EL, Rogers NL, Cizman Z, Goldenstein SK, Vogler C, Venkataraman S, Metaxas DN. ***Optical computer recognition of facial expressions associated with stress induced by performance demands.*** **Aviat Space Environ Med 2005; 76(6, Suppl.):B172–82.**

Application of computer vision to track changes in human facial expressions during long-duration spaceflight may be a useful way to unobtrusively detect the presence of stress during critical operations. To develop such an approach, we applied optical computer recognition (OCR) algorithms for detecting facial changes during performance while people experienced both low- and high-stressor performance demands. Workload and social feedback were used to vary performance stress in 60 healthy adults (29 men, 31 women; mean age 30 yr). High-stressor scenarios involved more difficult performance tasks, negative social feedback, and greater time pressure relative to low workload scenarios. Stress reactions were tracked using self-report ratings, salivary cortisol, and heart rate. Subjects also completed personality, mood, and alexithymia questionnaires. To bootstrap development of the OCR algorithm, we had a human observer, blind to stressor condition, identify the expressive elements of the face of people undergoing high- vs. low-stressor performance. Different sets of videos of subjects' faces during performance conditions were used for OCR algorithm training. Subjective ratings of stress, task difficulty, effort required, frustration, and negative mood were significantly increased during high-stressor performance bouts relative to low-stressor bouts (all $p < 0.01$). The OCR algorithm was refined to provide robust 3-d tracking of facial expressions during head movement. Movements of eyebrows and asymmetries in the mouth were extracted. These parameters are being used in a Hidden Markov model to identify high- and low-stressor conditions. Preliminary results suggest that an OCR algorithm using mouth and eyebrow regions has the potential to discriminate high- from low-stressor performance bouts in 75–88% of subjects. The validity of the workload paradigm to induce differential levels of stress in facial expressions was established. The paradigm also provided the basic stress-related facial expressions required to establish a prototypical OCR algorithm to detect such changes. Efforts are underway to further improve the OCR algorithm by adding facial touching and automating application of the deformable masks and OCR algorithms to video footage of the moving faces as a prelude to blind validation of the automated approach.

Keywords: optical computer recognition, computer vision, workload, performance, stress, human face, cortisol, heart rate, astronauts, Markov models.

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 can and have occurred during long-duration missions, especially when operational performance demands become elevated when unexpected and/or underestimated operational requirements occurred while crews were already experiencing work-related stressors (13,28,42,43,52,57,66). In some of these instances, stressed flight crews have withdrawn from voice communications with ground controllers (7,66), or when pressed to continue performing, made errors that could have jeopardized the mission (13,28). 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 flight to other locations in the solar system.

Facial Expressions of Stress

Measurement of human emotional expressions via the face, including negative affect and distress, dates back to Darwin (14), but in recent years has been undergoing extensive scientific study (46). Although cultural differences can intensify facial expression of emotions (53), there is considerable scientific evidence that select emotions are communicated in distinct facial displays across cultures, age, and gender (45). Because many techniques for monitoring stress reactions are impractical, unreliable, or obtrusive in spaceflight, we seek to develop a novel, objective, unobtrusive computer vision system to continuously track facial expressions during performance demands, to detect when

From the Unit for Experimental Psychiatry, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA (D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman); and the Center for Computational Biomedicine, Imaging and Modeling, Department of Computer Science, Rutgers University, New Brunswick, NJ (S. K. Goldstein, C. Vogler, S. Venkataraman, D. N. Metaxas).

Address reprint requests to: David F. Dinges, Ph.D., Professor and Director, Unit for Experimental Psychiatry, Department of Psychiatry, University of Pennsylvania School of Medicine, 1013 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021; dinges@mail.med.upenn.edu.

performance-impairing stress reactions may be present during work.

Measurements of specific facial expressions as correlates of emotional states have historically been based on either electromyographic (EMG) recordings of facial muscles (45,47,60,70), or temporally static 2-dimensional pictures scored by human observers using standardized criteria such as the facial affect coding system (FACS) developed by Ekman (25–27). Both EMG and the FACS can be used to identify specific prototypical emotional states (e.g., surprise, sadness, fear, anger) with a relatively high degree of accuracy (27). However, despite their precision and accuracy, neither technique is practical for use in spaceflight. Facial EMG is an intrusive measure, and the FACS is impractical due to its reliance on human observation. An online, optically based computerized recognition system for measuring behavioral stress via facial expressions may be able to overcome the practical limitations of these techniques. To help identify the development of neurobehavioral dysfunction before it becomes clinically or operationally significant, an optical computer recognition (OCR) system needs to be a valid, reliable, and practical monitor of stress reactions expressed in the face. It must also have capabilities beyond current computer-based systems for facial recognition (31,41,49,65,78).

OCR of Facial Expressions

Existing approaches to optical tracking and recognition of faces can be classified into two-dimensional and three-dimensional methods. Two-dimensional methods typically seek to recognize the characteristics of the face directly from two-dimensional decompositions and transformations of the image (2,48,73). The primary drawback of these approaches is that they are not sensitive to rotations and translations of the face. As a result, there are tight constraints on the posture of the face with respect to the camera which are not acceptable for monitoring humans at work. In contrast, three-dimensional methods typically use deformable models to track the face. These models yield a three-dimensional parameterization of the face—either of coordinates of points on the face, or of more complex building blocks, such as the shape of the cheeks. These parameterizations can then be used in the OCR algorithm. Existing tracking methods use edge-based features or optical flow as input from the images. When both optical flow and deformable models have been used, the computation of the optical flow and the model's deformations have not been coupled (30,31,36,51), which leads to inaccuracies in tracking. To overcome these limitations, Metaxas and colleagues (15–18) developed a formal framework for the integration of edges and optical flow within a deformable model framework and applied it to facial shape and motion estimation. This method tracks the three-dimensional shape of the face from one single camera with a deformable model, while coupling it with optical flow in the form of a constraint.

Tracking of the face alone, however, is not enough. An effective OCR system must be capable of recognizing facial expressions from the results of the tracking. Bartlett and colleagues (2) have been able to recognize relatively subtle facial expression units using a two-dimensional-based hybrid between neural networks and optical flow-based computations within the limitations of two-dimensional-based approaches. Essa and Pentland have successfully recognized some of the classical expressions, such as surprise and anger (30,31), but their approach is limited by the accuracy of their three-dimensional (3D) tracking methods. The 3D shape estimation and motion tracking of articulated objects in general is a difficult research problem. The main difficulties stem from the complex 3D motions of humans and occlusion. There have been a series of efforts to model, estimate, and track humans and their body parts (4,5,19,31,33,37–40,48,50,58,59,61,67,74–77). The success of these methods depends largely on the proper integration of the visual cues used for shape and motion estimation.

Our optical computer recognition approach uses deformable model theory and statistical methods for the estimation of human facial expressions and motions. Deformable models have been used in a variety of areas and applications in computer vision for tracking and shape estimation (19,32,63). Most of these approaches have been deterministic; that is, they did not address the statistical uncertainties inherent in tracking images and fitting models to a particular shape or image. Statistical methods to address these issues have recently been introduced and explored, but these methods do not scale well with the number of parameters to be estimated or else make assumptions about the shape and the characteristics of the probability distribution functions. Generally, little is known about these functions except their bounds. We have used deformable model theory and augmented it with novel algorithms for detection of human expressions, as well as stochastic methods based on Affine arithmetic to embed deformable models within a statistical framework. This method was developed specifically to track faces (34). Unlike many previous statistical approaches in computer vision, this approach avoids making assumptions about the probability distribution functions, and scales well with the number of parameters used in the deformable model description.

Rather than attempting to identify specific emotional states, our OCR system was devised to identify facial expressions evoked in response to elevated performance demands (i.e., a behavioral stressor condition). It is possible that such reactions may be individually unique and comprised of different expressions of negative affect. Therefore, we believe it is more appropriate to assess these expressions relative to subjective impressions of stress and frustration rather than to assume correspondence to a specific affective expression (e.g., sadness, anger, or fear). Once the validity and reliability of the OCR system is established, we can address whether there is merit to training the system further to detect specific affective states. In this paper we describe the development of our optical computer recognition approach for recognizing facial reactions to stress induced by variations in cognitive performance demands.

METHODS

An experiment designed to vary stress reactions using performance demands was conducted on 60 healthy adults (29 men, 31 women; 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 (3) were screened out of the study. Eyewear was not allowed (except for contact lenses), as glasses obscure 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.

Procedures

During the in-laboratory session, subjects completed questionnaires measuring aspects of personality, mood, stress perception, and coping strategies to determine if variations on these psychological parameters related to the identification of facial expressions of stress. These scales included: the Toronto Alexithymia Scale (71); State/Trait Anxiety Index (68); Beck Depression Inventory (3); Perceived Stress Scale (10); Life Orientation Test (64); The COPE Inventory (9); Distress Inventory; Morningness/Eveningness Questionnaire (35); Millon Index of Personality Styles (56); Rotter's Locus of Control (62); Pittsburgh Sleep Quality Index (8); Marlowe-Crowne Social Desirability Scale (12); and Penn State Worry Questionnaire (55).

On 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 (22,24). 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 analogue ratings and mood scores, salivary cortisol, and heart rate.

Subjects were studied twice during both low and high workload performance demands—once to provide data to train the OCR algorithm, and once to provide data to "test" the algorithm. High workload scenarios induced greater stress by involving more difficult performance tasks, negative social feedback, and greater time pressure relative to low workload scenarios. In addition, stress was further increased by social feedback. Periodically throughout low and high workload test bouts, on-screen feedback was provided. During the low workload periods, this feedback was positive (i.e., low-stressor scenario). During the high workload periods, feedback was negative (i.e., high-stressor scenario). Social 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, salivary cortisol, and heart rate (the physiological data will be reported elsewhere).

Workload tasks: The following tasks were administered to subjects in an easy version during the low workload test bouts and a more difficult version during the high workload test bouts: the Stroop word-color interference task (69), the psychomotor vigilance task (23), the probed recall memory (20), the descending subtraction task (21), the visual memory task, the digit symbol substitution task, the serial addition subtraction task (72), the synthetic workload task (29), the meter reading task, the logical reasoning task (1), and the Haylings sentence completion task (6). High workload versions of the tasks required more attention and skill and faster responding.

Subjective measures of mood and stress: Before and after every workload bout subjects responded to visual analogue scales for alertness, mental and physical fatigue, exhaustion, and stress. Following each workload bout, subjects completed the Profile of Mood States (POMS), an adjective checklist for assessing six dimensions of mood (fatigue-inertia, vigor-activity, confusion-bewilderment, tension-anxiety, anger-hostility, depression-dejection) and overall general mood disturbance (54). Following every task, subjects completed visual analogue scales to rate distress, effort, frustration, and difficulty of tasks along with pre- and post-test stress questionnaires assessing subjects' perception of how stressed they felt at that point in time.

Human Recognition of Stress Expressed in the Face

Videos of subjects' faces were made throughout both low-stressor and high-stressor scenarios. They were blind coded for analyses without audio information. To bootstrap development of the OCR algorithm by providing information on facial expressions that discriminated low- and high-stressor conditions, we had a human observer blind to stressor condition attempt to identify the expressive elements of the face associated with subjects undergoing high- vs. low-stressor performance scenarios. The human scorer used a facial scoring tool based in part on the facial affect coding system of Ekman (25–27) and correctly classified (as low- vs. high-stressor) 85% of facial videos in the 60 subjects. This is well above chance, and indicates the minimum standard against which to judge the OCR algorithm of facial expressions being developed. We compared human scorer-identified facial changes when accurate vs. inaccurate. Accuracy was found to be higher when subjects showed more left eyebrow movements, biting of the lower lip, simultaneous lateral movements of the lips, and more total body movements in high workload conditions than during low workload conditions (all $p < 0.05$). For faces correctly classified by the human observer into high-stressor vs. low-stressor performance conditions, these human-scored aspects of facial expressions served to help inform (bootstrap) the development of the discrimination algorithm for the optical computer algorithm.

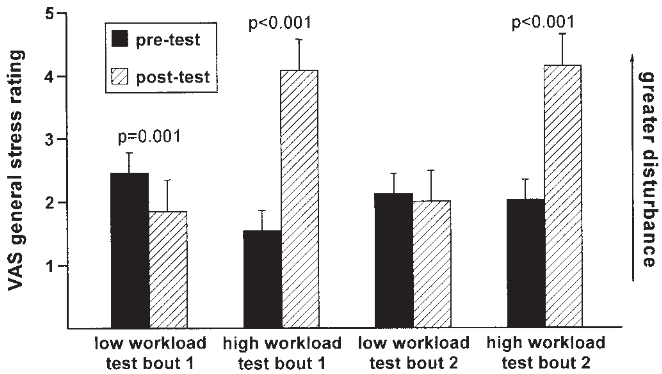

Fig. 1. Mean ± SEM from n = 60 healthy adults who rated the stress level on a visual analogue scale completed prior to (solid bars) and immediately following (hatched bars) each of four performance workload test bouts. A higher score represents greater general stress. There was a significant decrease in stress ratings immediately after the first (low workload) bout ($p = 0.001$). For both of the high workload test bouts (1 and 2) there were significant increases in subjective ratings of general stress from the beginning to the end of the bouts ($p < 0.001$).

Development of the Optical Computer Recognition Algorithm and Deformable Masks

Significant advances in programming and enhancement of the capabilities of the computer recognition algorithm were completed to develop a technique for robust 3D tracking of facial expression. Using this approach, numerous low-level, well-known computer vision algorithms were initially used to extract two-dimensional (2D) information (tracking single points, edges, shading, and optical flow) from a video-recording of the face. A newly developed statistical technique, referred to as cue integration, uses principles of Affine arithmetic and Gaussian distribution mathematics to combine information from these low-level algorithms using a maximum likelihood estimator to perform optimal data fusion. The integrated information is used to construct a deformable model or 3D surface representation of the face. Put simply, the new technique, developed by Metaxas and colleagues (34), allowed the translation of 2D video footage to a form that could be used for tracking the 3D orientation and translation of the face, as well as parameters that described the movement of eyebrows, mouth, etc. This novel technology provided an unobtrusive mechanism for the measurement of changes in facial expression and formed the initial OCR algorithm detection of facial changes.

RESULTS

Performance Workload and Stress Ratings

Fig. 1 reveals that subjects' ratings of general stress level during performance tests increased significantly following completion of high workload test bouts compared with low workload test bouts. Virtually identical profiles were found for ratings of difficulty of tasks, effort required to perform, and frustration (**Fig. 2**). In all cases high workload performance demands significantly increased these negative experiences relative to low workload demands. In addition, high workload demands induced a significant elevation in POMS total mood disturbance scores (**Fig. 3**), as well as significantly more negative moods reports for all POMS subscales (tension-anxiety, anger-hostility, depression-dejection, confusion-bewilderment, fatigue-inertia, vigor-activity; all $p < 0.01$). As expected, subjects performed significantly worse on the performance tasks in the high workload (high stressor) bout than they did in the low workload (low stressor) bout (all $p < 0.01$).

Alexithymia

A concern in development of OCR of facial expressions of stress in spaceflight was the belief that astronauts may (learn to) be alexithymic and hence would not express feelings of distress in their faces. Persons who are alexithymic have difficulty identifying and describing their feelings, difficulty distinguishing between their feelings and bodily sensations, and a preoccupation with external events. We assessed whether there was a relationship between alexithymia and the subjective ratings of stress induced by high workload demands. The distribution of alexithymia scores across our sample of 60 healthy adults revealed that although the majority fell within the normal range, approximately 25% were classified as mild to moderately alexithymic (i.e., score of 63 to 80; see **Fig. 4**). Consistent with the scientific literature (44,71), alexithymia scores were positively correlated with ratings of subjective stress during the high workload bouts ($\rho = 0.34$, $p = 0.011$) and with total mood disturbance on the POMS ($\rho = 0.33$, $p = 0.021$), but not with ratings of task difficulty, effort required, or frustration. Future studies will evaluate the accuracy of the OCR algorithm of facial stress as a function of alexithymia and the effects of microgravity on facial edema.

Optical Computer Recognition of Stress Expressed in the Face

Static, two-dimensional facial expressions from two subjects are shown in **Fig. 5**, illustrating some of the performance-induced expressions seen in the experiment. **Fig. 6** illustrates how the optical computer recognition algorithm has evolved to significantly increase the quality, properties, and precision of the deformable face mask relative to facial areas of interest in stress expression (e.g., eyebrows, forehead, mouth). In the first generation of the mask (Fig. 6A) we used a simple discretization of the face, with only 192 nodes. The model had seven parameters to describe the rigid transformation of the face, and one to describe eyebrow movement. Although simple and easy to handle, this mask did not capture enough detail of the shape and texture of the face to allow good tracking of facial features. In the second generation (Fig. 6B), we started to employ a more refined discretization of the face at rest, with 1100 nodes. Just as in the first generation, the model had seven parameters to describe the rigid transformation and one for eyebrow movement. With this version of the mask we achieved tracking of a small number of sequences; however, the mask was unable to maintain track of the face whenever movement of the mouth occurred. In the third, and current, generation (Fig. 6C), we added three mouth deformation parame-

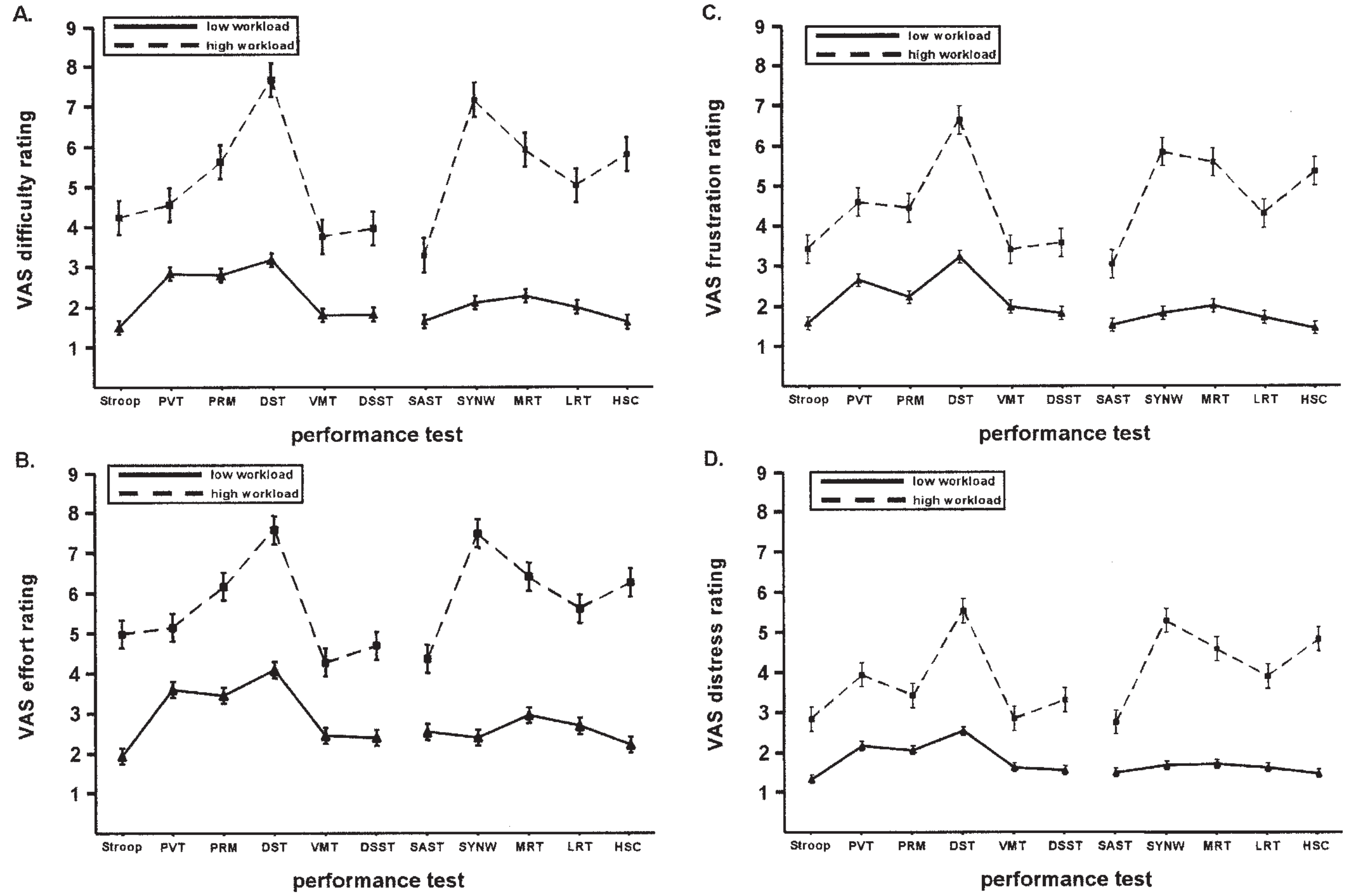

Fig. 2. Mean ± SEM from n = 60 healthy adults who completed rated difficulty of the tasks just completed (panel A); effort to perform the tasks just completed (panel B); frustration level during the tasks just completed (panel C); and distress level following the tasks just completed (panel D). Higher scores represent greater negative experiences. Subjective ratings of difficulty, effort, frustration and distress were significantly higher following all tasks in the high workload test bouts, relative to the low workload test bouts (all $p < 0.01$). For each performance task, an easy version was used in the low workload test bouts, and a difficult version was used in the high workload test bout. Task abbreviations: psychomotor vigilance task (PVT); probed recall memory (PRM); descending subtraction task (DST); visual memory task (VMT); digit symbol substitution task (DSST); serial addition/subtraction task (SAST); synthetic workload task (SYNW); meter reading task (MRT); logical reasoning task (LRT); Haylings sentence completion (HSC).

ters to the mask, simulating the risorius and zigomatic major muscles, and the jaw opening. We created a semi-automatic mechanism to "fit" the generic mask to any particular individual. With this model, and the new tracking techniques we have developed, we are now able to track over long sequences (thousands of frames).

Computerized Recognition of Stress Expressed in the Face

The computational detection of whether a subject is undergoing a stress reaction in response to a high workload can be decomposed into two stages. The first stage consists of recognizing the possible individual displays of a stress response in the human face, such as eye movements and blinking, and various negative facial expressions. The second stage consists of accumulating the information collected in the first stage, and deciding whether these displays occur frequently enough to classify as a stress response.

First stage: Recognition of individual stress-related displays: From an abstract point of view, the first stage corresponds to the task of detecting specific patterns in a time-varying data signal. Depending on the specific task and the pattern that we are looking for, the signal

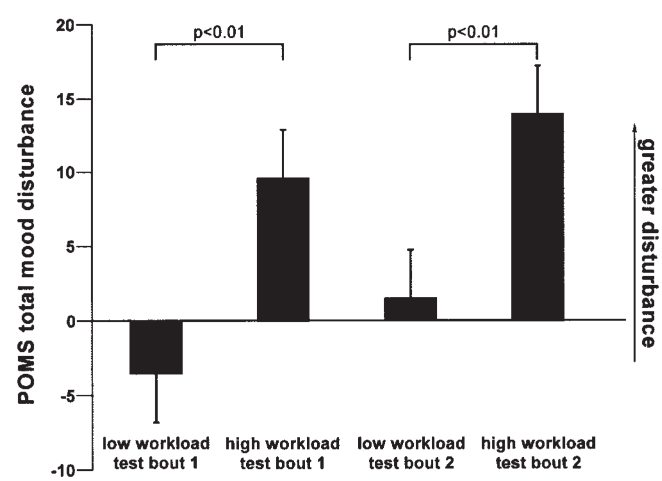

Fig. 3. Changes in Profile of Mood States (POMS) total mood disturbance scores (mean ± SEM) from before to immediately after each performance test bout for n = 60 healthy adults. Total mood disturbance scores are calculated from the scores of each of the subscales of the POMS: anger/hostility, confusion/bewilderment, tension/anxiety, fatigue/inertia and vigor/activity, and then standardized to population norms. Higher positive scores indicate a greater increase in negative mood states as a function of performance. For both of the high workload test bouts, total mood disturbance was significantly elevated compared with the low workload test bouts ($p < 0.01$).

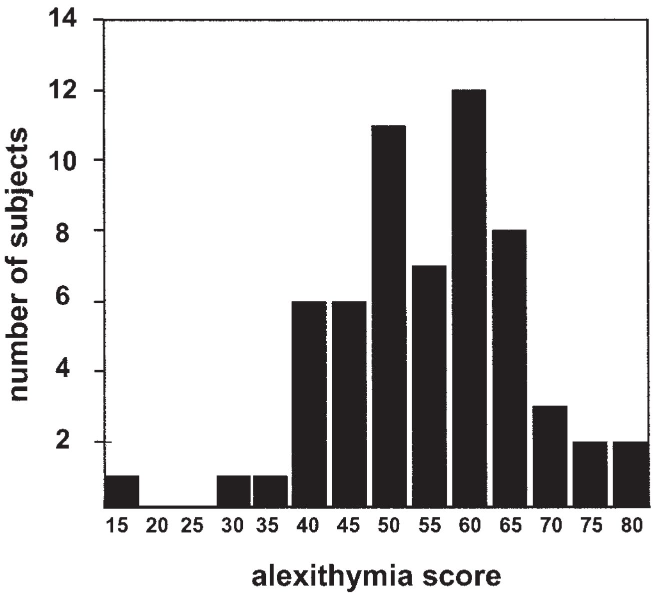

Fig. 4. Distribution of alexithymia scores, as measured by the Toronto Alexithymia Scale (54), in n = 60 healthy adults completing both low and high workload performance tests. Following the literature, scores of 63 or greater were classified as alexithymic, which made up 25% of the study sample.

often simply consists of the deformable model parameter vectors that we estimate during the face tracking process. For instance, in order to detect rapid head movements, we are interested in the rigid body component of the parameter vector. The orientation, position, and derivatives thereof contain all the necessary information for this particular pattern. Likewise, in order to detect specific negative facial expressions, we are interested in the non-rigid deformation component of the parameter vector, which controls the eyebrow, mouth movements, and nearby regions. Eye blinking is slightly more complicated to handle because the deformable model does not contain any parameters that control it directly. However, output from the tracking algorithm does contain information on the location of the eyes in the human face at any given point in time. Eyes are represented as holes in the deformable model, delimited by the region that is formed by a set of nodes. Therefore, from these deformable model parameters, the position of these nodes can be deduced, making it possible to project their positions into image space and find out which region in the video frame corresponds to the eyes. We then use a grayscale level averaging method on the region in the video frame to determine the degree to which the eyes are opened or closed. Uniform grayscale levels indicate that the eyelids are covering the irises, whereas more diverse grayscale levels indicate that the irises are visible. Just like the model parameters, the degree of openness of the eyes can thus be quantified in a few numbers, which is important for the subsequent training of the recognition algorithm.

Detecting facial stress patterns in a time-varying signal is very similar to other well-known activity recognition tasks, such as gesture recognition, and sign language recognition. For these tasks, hidden Markov models (HMMs) have been shown to be highly suitable for several reasons: 1) the HMM recognition algorithm (Viterbi decoding) is able to segment the data signal into its constituent components implicitly, so it is not necessary to concern ourselves with the often extremely difficult problem of segmenting a data signal explicitly; 2) the state-based nature of HMMs is a natural match for the task of recognizing signals over a period of time; and 3) the statistical nature of HMMs makes them ideal

Fig. 5. Examples of facial images from subjects performing on the low workload test battery (frames A1, B1) and while performing on the high workload test battery (frames A2–4, B2–4). Like most subjects evaluated, facial expressions tend to be relatively expressionless during low workload performance, but episodically more expressive during high workload performance, which was rated as more stressful (Fig. 1), more difficult (Fig. 2), more frustrating (Fig. 2), and more negative mood states (Fig. 3). Subjects provided written permission for display of facial images.

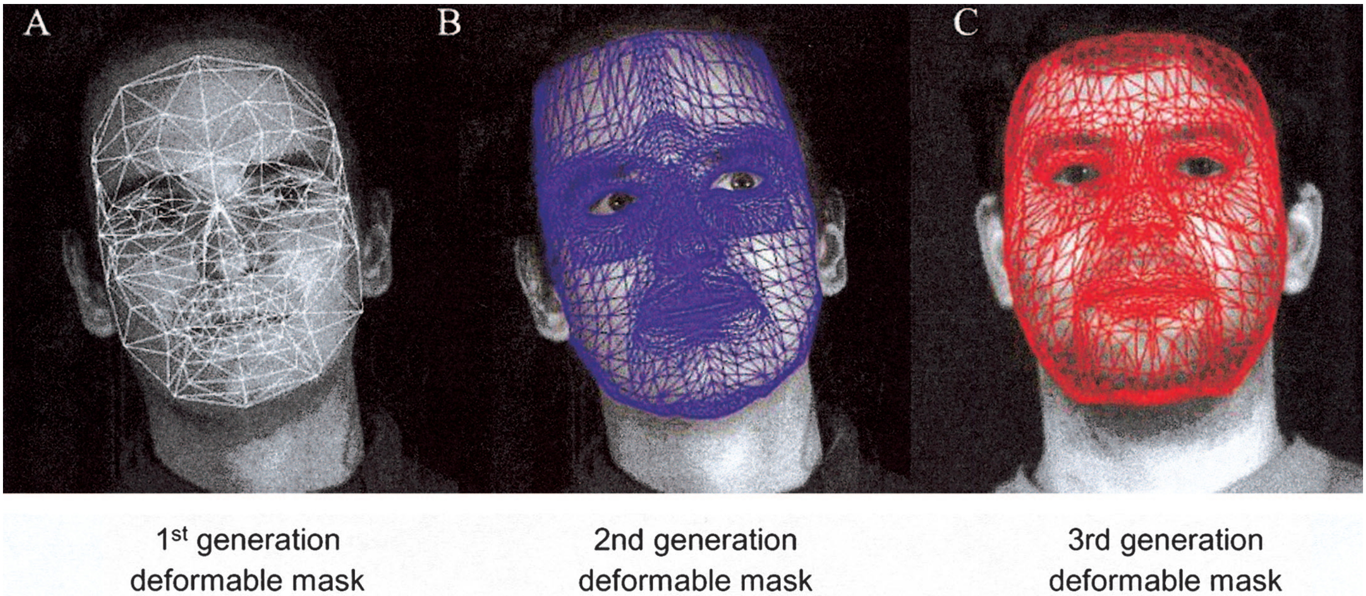

Fig. 6. Progressive improvements in programming of the optical computer algorithm mask have increased the quality, properties, and precision of the deformable face masks. Panel C, the third generation mask, is the one currently being used in the determination of stress in subjects undergoing the low vs. high workload protocol.

for recognizing tasks that exhibit an inherent degree of variation (for example, due to motor limitations, humans generally do not perform the same movement twice in exactly the same way, even if they intend to). To make HMM-based recognition of potentially stress-related displays work, we first train the HMMs on hand-labeled examples of such displays. The labeled examples include information on the starting and ending frame of the display, as well as the class into which it belongs (a specific type of negative facial expression, rapid head movement, eye blinking) and so on. Then, during the recognition phase, the HMM algorithm detects from the tracking data which of these types of displays occur in the video, and when (i.e., at which frames).

Second stage: Accumulation of the information on stress-related displays: The HMM-based recognition algorithm from the first stage delivers information on how often specific patterns occur in the human face, for how long, and when. Unlike in the first stage, there is no clear time-varying pattern that can be recognized. Instead, the presence of stress is most likely indicated by how often these patterns occur, and especially by the manner in which multiple patterns occur simultaneously, such as eye blinking together with rapid head movements. Thus, it is important not only to count the frequency of patterns in isolation, but also to determine how they are correlated. Since we have now developed the deformable tools, we are in a position to conduct many tracking analyses of the results and tune our models to recognize the various workload stress levels. Handling and training the possible correlations in a tractable manner is a difficult problem because the dimension of the space in which all these possible patterns can co-occur is too large to be captured computationally. Instead, an attempt to isolate causes and effects via decomposition into a Bayesian network may be more feasible. Basically, the Bayesian network needs to be trained, such that we end up with a probability that, given a certain number of co-occurrences of patterns, it is likely stress is present.

We use the fitted shape and the tracking of the deformable model to estimate the rigid body parameters of the face, as well as eyebrow and lip deformations. At this stage, the model is not yet powerful enough to capture asymmetric lip deformations. For this reason, we take an image-based approach. We use the image being tracked as a texture map for the deformable model. We know the 3D coordinates of each point on the model, thanks to our estimate of the rigid body parameters from the tracking process. We project these into image space, which yields the texture coordinates for each face. After texturing the model, we then rotate and translate it into a canonical position with the face oriented straight forward. The results are images of the face at every frame which are free of most distortions that resulted from rotations and translations. Only the non-rigid distortions remain, such as the shape of the mouth, which we would like to use as features for recognition. In particular, asymmetric mouth movements and shapes are a promising indicator for stress recognition. We compute the approximate contour of the lips for the face images by using an edge detector. If we then scan the contour of the lips along the x-axis, we can compute the distance between the upper and lower lips for each point on the lips. The edges are inherently noisy, consequently the distances are both noisy and unreliable. However, in aggregate form edges yield valuable information, and while the means and covariances of the distances do not contain enough information, a plot of the lip distances does. To bring these curves into a form that the computer can recognize, we fit a polynomial to them—in this case a simple line suffices. The slope of the line indicates the degree of symmetry, whereas the intercept strongly indicates how the lips behave overall in relation to the nose and

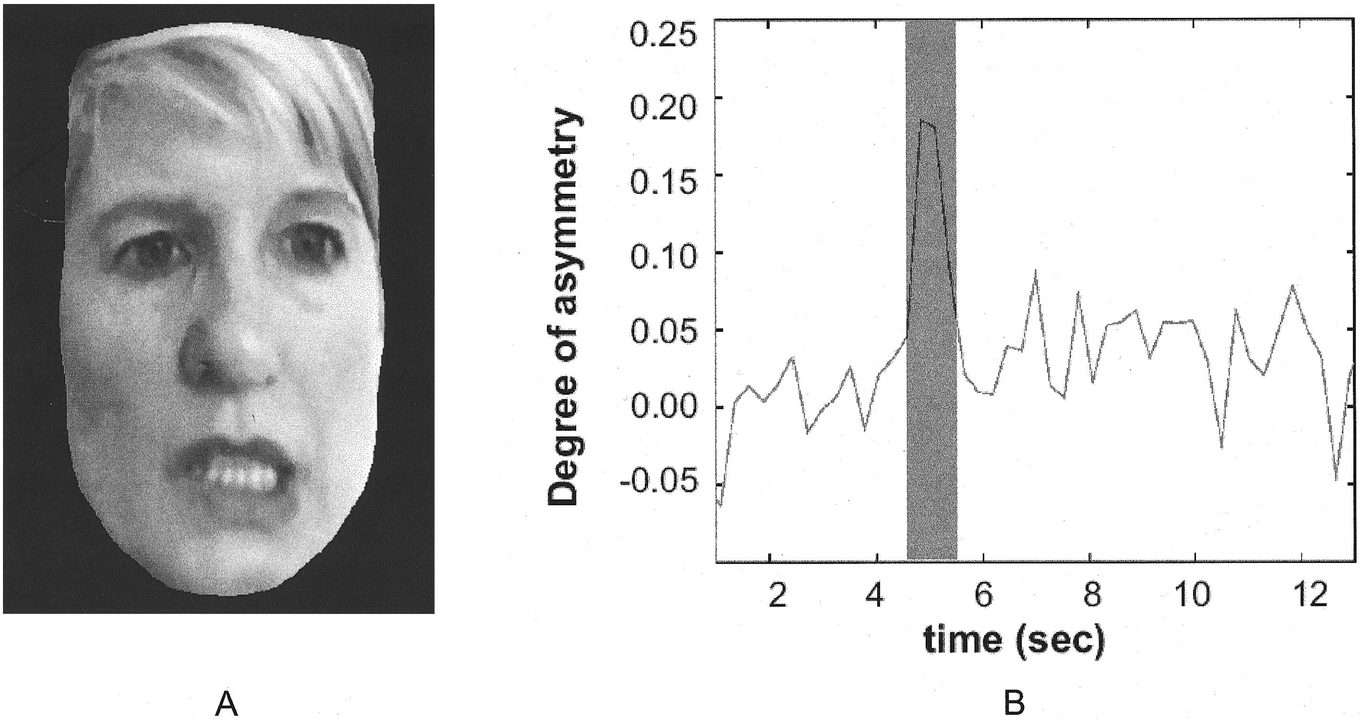

Fig. 7. Example output from the optical computer algorithm. The photo on the left (A) shows a facial expression during high workload (stressor condition). The graph on the right (B) is a representation of lip movement displayed by the subject. It quantifies lip distance plots into the slopes of the best-fit 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 region (gray bar in B) corresponds to the stressed expression in picture A. Subject provided written permission for display of facial image.

the rest of the face. **Fig. 7** shows how the slope of the line evolves over time (Fig. 7B) relative to the facial expression (Fig. 7A) under high workload. The highlighted region (gray bar in B) corresponds to the stressed expression in picture A. As the spike in the graph indicates, the slope strongly stands out for this expression and is easily recognizable by our computer algorithm.

Preliminary OCR accuracy results: Facial mouth and lip movements (e.g., asymmetries) and eyebrow movements, which the human scorer found could discriminate low- from high-stressor workload conditions, were use to train the OCR algorithm. The algorithm was then applied to low-stressor vs. high-stressor video from the first 20 subjects. The OCR algorithm correctly discriminated low- from high-stessor conditions in 15 of the 20 subjects (75%). However, data from three subjects were not reliable due to facial lighting that interfered with stable tracking of facial features. When these subjects were excluded from consideration, the OCR algorithm corrrectly discriminated 15 of the remaining 17 subjects (88%). Consequently, preliminary results suggest that an OCR algorithm using mouth and eyebrow regions has the potential to discriminate high- from low-stressor performance bouts in 75–88% of subjects. These will need to be confirmed on the remaining 40 subjects.

DISCUSSION

This project seeks to develop and test computer vision of facial expressions as an objective and unobtrusive measure of the presence of stress induced by workload performance demands. Substantial progress was made toward this goal. Performance workload was used to induce feelings of stress in subjects under high workload conditions (relative to low workload conditions). We were successful in experimentally increasing subjective stress experienced via workload demands. Ratings of general stress increased significantly following the completion of the high workload test bouts compared with the low workload test bouts. Virtually identical profiles were found for ratings of difficulty of tasks, effort required, and frustration. High workload performance demands significantly increased these ratings relative to low workload demands. High workload demands also induced a significant elevation in total mood disturbance scores, as well as significant effects in the direction of more negative moods reports for all mood subscales.

One potential barrier to developing optical computer recognition of human facial expressions of stress is alexithymia, which refers to having difficulty identifying and describing one's feelings, difficulty distinguishing between feelings and bodily sensations, and a preoccupation with external events. We found that alexithymia was related to stress responses and it will have to be more fully evaluated relative to OCR of facial expressions.

Substantial progress was made on increasing the sophistication of the OCR system to detect facial changes during performance stress. The final laborious phase of the project, involving the application of the deformable masks and optical computer algorithm to video images of the subjects while performing in the low- vs. high-workload conditions, is still ongoing. However, significant advances in programming and enhancement of the capabilities of the computer recognition algorithm have been completed to develop a technique for robust 3D tracking of facial expression. Using this approach, numerous low-level, well-known computer vision algorithms were initially used to extract 2D information (tracking single points, edges, shading, and optical flow) from video-recording of each subject's face during low and high workload performance demands. A newly developed statistical technique, referred to as cue integration, was used (based on principles of Affine arithmetic and Gaussian distribution mathematics to combine the information from these low-level algorithms using a maximum likelihood estimator) to perform optimal data fusion. The integrated information was used to construct a deformable model, or 3D surface presentation of the face. This approach allowed us to avoid making assumptions about the probability distribution functions, unlike most previous statistical approaches in computer vision, and scales well with the number of parameters used in the deformable model description. Put simply, the new technique, developed by Metaxas and colleagues, allowed the translation of 2D video footage to a form that could be used for tracking the 3D orientation and translation of the face, as well as parameters that describe the movement of eyebrows, mouth, etc. This novel technology provided us with a sensitive and unobtrusive mechanism for the measurement of changes in facial expression. Preliminary results suggest that an OCR algorithm using mouth and eyebrow regions has the potential to discriminate high- from low-stressor performance bouts in at least 75–88% of subjects. We believe we can replicate and improve on this degree of accuracy by further development of the OCR algorithm.

We are continuing with the development of the deformable face masks to increase the quality and precision of the tracking and the ability to accurately detect changes in facial features that are representative of distressed states. The next stages will have asymmetric deformations, extra degrees of freedom for the eyebrows (decoupling frontalis from corrugator), and better lip movement description. We plan to allow automatic multi-resolution selection for the discretization of the mask, using as many or as few points as it is necessary as well as automated face detection and mask initialization algorithms. Finally, the OCR algorithm will be expanded to include touching the face (e.g., ear poking and pulling, nose and beard scratching, eye-wiping, etc.). Face touching has been reported to be associated with stress and tension (11), and it occurred not infrequently in some of our subjects undergoing high-stressor conditions (e.g., see Fig. 5B3).

The knowledge gained has the potential to identify an objective, unobtrusive, automated method for the recognition, monitoring, and management of the risks of neurobehavioral dysfunction due to work-related stress in spaceflight and in many Earth-based safety-sensitive occupations such as transportation workers (e.g., truck drivers, train conductors, airline pilots), operators in safety-sensitive industries (e.g., power plant control rooms), and military personnel. Validation of this computer vision system will provide a critically needed unobtrusive, objective method for detecting the development of stress responses in astronauts, and it will form a key component in the prevention and countermeasure strategies for stress in long-duration spaceflight.

ACKNOWLEDGMENTS

This research was supported by NASA cooperative agreement NCC 5–98 with the National Space Biomedical Research Institute. We thank Adrian Ecker, John W. Powell IV, Michele Carlin, Cristos Ballas, and Olivera Crenshaw for making the project technically feasible. We acknowledge the suggestion of Professor George G. Coss that facial touching may provide an additional component in an OCR algorithm for discriminating stress states.

REFERENCES

1. Baddeley A. A 3-min reasoning task based on grammatical transformation. Psychonomic Science 1968; 10:341–2.
2. Bartlett MS, Hager JC, Ekman P, Sejnowski TJ. Measuring facial expressions by computer image analysis. Psychophysiology 1999; 36:253–63.
3. Beck AT, Steer RA. Internal consistencies of the original and revised Beck Depression Inventory. J Clin Psychol 1984; 40: 1365–7.
4. Bharatkumar AG, Daigle KE, Pandy MF, et al. Lower limb kinematics of human walking with the medial axis transformation. Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects; 11–12 Nov 1994; Austin, TX. New York: IEEE; 1994.
5. Bregler C, Malik J. (1998) Tracking people with twists and exponential maps. In: Proceedings of the IEEE Computer Society on Computer Vision and Pattern Recognition CVPR'98; 23–25 January 1998; Santa Barbara, CA. Washington, DC: IEEE Computer Society; 1998.
6. Burgess PW, Shallice T. Response suppression, initiation and strategy use following frontal lobe lesions. Neuropsychologia 1996; 34:263–73.
7. Burrough B. Dragonfly: NASA and the crisis aboard Mir. New York: Harper Collins; 1998.
8. Buysse DJ, Reynolds CF 3rd, Monk TH, et al. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989; 28:193–213.
9. Carver CS, Scheier MF, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989; 56:267–83.
10. Cohen S, Williamson GM. Perceived stress in a probability sample of the United States. Beverly Hills, CA: Sage; 1988.
11. Coss RG. The cut-off hypothesis: its relevance to the design of public places. Man-Environment Systems 1973; 3:417–40.
12. Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. J Consult Psychol 1960; 24:349–54.
13. Culbertson F. Testimony before the Committee on Science, U.S. House of Representatives; September 1997; 18:1997.
14. Darwin C. The expression of emotions in man and animals. New York: Oxford University Press; 1872/1998.
15. DeCarlo D, Meataxas D, Stone M. An anthropometric face model using variational techniques. Proceedings of ACM SIGGRAPH; 19–24 Jul 1998; Orlando, FL. New York: ACM; 1998.
16. DeCarlo D, Metaxas D. The integration of optical flow and deformable models with applications to human face shape and motion estimation. Proceedings of the IEEE Computer Society on Computer Vision and Pattern Recognition, CVPR'96; 18–20 June 1996; San Francisco, CA. Washington, DC: IEEE Computer Society; 1996.
17. DeCarlo D, Metaxas D. Deformable model-based shape and motion analysis from images using motion residual error. Proceedings of the IEEE Sixth International Conference on Computer Vision; 4–7 Jan 1998; Bombay, India. New York: IEEE; 1998.

18. DeCarlo D, Metaxas D. Combining information using hard constraints. Proceedings of the IEEE Computer Society on Computer Vision and Pattern Recognition, CVPR'99; 23–25 January 1999; Ft. Collins, CO. Washington, DC: IEEE Computer Society; 1999.
19. DeCarlo D, Metaxas, D. Optical flow constraints on deformable models with applications to face tracking. International Journal of Computer Vision 2000; 32:99–127.
20. Dinges DF, Kribbs NB, Bates BL, Carlin MM. A very brief probed-recall memory task: Sensitivity to sleep loss. Sleep Res 1993; 22:330.
21. Dinges DF, Orne EC, Evans FJ, Orne MT. Performance after naps in sleep conducive and alerting environments. In: Johnson LC, Tepas DI, Colquhoun WP, Colligan MJ, eds. Biological rhythms, sleep and shift work. Advances in sleep research. New York: Spectrum Publications 1981; 7:539–52.
22. Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk TH, ed. Sleep, sleepiness and performance. Chichester: John Wiley & Sons; 1991: 97–128.
23. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Methods Instrum Comput 1985; 17:652–5.
24. Dinges DF. Probing the limits of functional capability: the effects of sleep loss on short-duration tasks. In: Broughton RJ, Ogilvie RD, eds. Sleep, arousal, and performance. Boston: Birkhäuser; 1992:177–88.
25. Ekman P, Davidson RJ, Friesen WV. The Duchenne smile: emotional expression and brain physiology II. J Pers Soc Psychol 1990; 58:342–53.
26. Ekman P, Friesen W. Facial action coding system (investigator's guide). Palo Alto, CA: Consulting Psychologists Press, Inc.; 1978.
27. Ekman P. Facial expressions of emotion: an old controversy and new findings. Philos Trans R Soc Lond B Biol Sci 1992; 335: 63–9.
28. Ellis SR. Collision in space. Ergon Des 2000; 8:4–9.
29. Elsmore TF. Synwork1: a PC-based tool for assessment of performance in a simulated work environment. Behav Res Methods Instrum Comput 1994; 26:421–6.
30. Essa I, Pentland AP. Facial expression recognition using a dynamic model and motion energy. Proceedings of the IEEE Fifth International Conference on Computer Vision, ICCV'95; 20–23 June 1995; Boston, MA. New York: IEEE; 1995.
31. Essa I, Pentland AP. Coding, analysis, interpretation and recognition of facial expressions. IEEE PAMI 1997; 19:757–63.
32. Fua PV, Leclerc YG. Object-centered surface reconstruction: combining multi-image stereo and shading. IJCV 1995; 16:35–56.
33. Gavrilla DM, Davis LS. Towards 3-D model-based tracking and recognition of human movement. Proceedings of the IEEE International Workshop on Face and Gesture Recognition; June 1994; Zurich, Switzerland. New York: IEEE; 1995.
34. Goldenstein S, Vogler C, Metaxas D. Statistical cue integration in DAG deformable models. IEEE Transactions on Pattern Analysis and Machine Intelligence 2003; 25:801–13.
35. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1976; 4:97–110.
36. Jebara T, Pentland AP. Parametrized structure from motion for 3D adaptive feedback tracking of faces. Proceedings of the IEEE Fifth International Conference on Computer Vision CVPR'97; 17–19 June 1997; San Juan, Puerto Rico. New York: IEEE; 1997.
37. Kakadiaris IA, Metaxas D. 3D human body model acquisition from multiple views. The IEEE Fifth International Conference on Computer Vision; 20–23 June 1995; Boston, MA. New York: IEEE; 1995.
38. Kakadiaris IA, Metaxas D. Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. The IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 18–20 June 1996; San Francisco, CA. Washington, DC: IEEE Computer Society; 1996.
39. Kakadiaris IA, Metaxas D, Bajcsy R. Inferring object structure in 2D from the deformation of apparent contours. Journal of Computer Vision and Image Understanding 1997; 65:129–47.
40. Kakadiaris IA, Metaxas D. Model based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000; 22:1453–9.
41. Kanade T, Cohn JF, Tian Y. Comprehensive database for facial expression analysis. The 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00); 28–30 Mar 2000; Grenoble, France. New York: IEEE; 2000.
42. Kanas N, Salnitskiy V, Grund E, et al. Social and cultural issues during Shuttle/MIR space missions. 50th International Astronautical Congress; 4–8 Oct 1999; Amsterdam, The Netherlands. Paris, France: IAF; 1999.
43. Kanas N. Psychiatric issues affecting long duration space missions. Aviat Space Environ Med 1998; 69:1211–6.
44. Kano M, Fukudo S, Gyoba J, et al. Specific brain processing of facial expressions in people with alexithymia: an H2 15O-PET study. Brain 2003; 126:1474–84.
45. Keltner D, Eckman P, Gonzaga GC, Beer J. Facial expression of emotion. In: Davidson RJ, Scherer KR, Goldsmith HH, eds. Handbook of affective sciences. New York: Oxford University Press; 2003:415–32.
46. Keltner D, Ekman P. Introduction: expression of emotion. In: Davidson RJ, Scherer KR, Goldsmith HH, eds. Handbook of affective sciences. New York: Oxford University Press; 2003: 411–4.
47. Lane RD, Reiman EM, Bradley MM, et al. Neuroanatomical correlates of pleasant and unpleasant emotion. Neuropsychologia 1997; 35:1437–44.
48. Lanitis A, Taylor CJ, Cootes TF. A unified approach for coding and interpreting face images. Proceedings of the IEEE Fifth International Conference on Computer Vision, ICCV '95; 20–23 June 1995; Boston, MA. New York: IEEE; 1995.
49. Lanitis A, Taylor CJ, Cootes TF. Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997; 19:743–56.
50. Leung MK, Yang YH. An empirical approach to human body analysis. Technical Report. Saskatoon, Saskatchewan, Canada: University of Sasketchewan; 1994.
51. Li H, Roivanen P, Forchheimer R. 3-D motion estimation in model-based facial image coding. IEEE Transactions on Pattern Analysis and Machine Intelligence 1993; 15:545–56.
52. Linenger JM. Off the planet. New York: McGraw Hill; 2000.
53. Marsh AA, Elfenbein HA, Ambady N. Nonverbal "accents": cultural differences in facial expressions of emotion. Psychol Sci 2003; 14:373–6.
54. McNair DM, Lorr M, Droppleman LF. EITS manual for the profile of mood states. San Diego: Educational and Industrial Testing Service; 1971.
55. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State worry questionnaire. Behav Res Ther 1990; 28:487–95.
56. Millon T. A theoretical derivation of pathological personalities. In: Millon T, Klerman G, eds. Contemporary directions in psychopathology. New York: Guilford; 1986.
57. Niller E. "The right stuff" may no longer be the best stuff for inhabitants of new space station. The Boston Globe; April 4, 2000:E4.
58. Rehg JM, Kanade T. Visual tracking of high DOF articulated structures: an application to human hand tracking. Third European Conference on Computer Vision; 2–6 May 1994; Stockholm, Sweden. New York: Springer-Verlag; 1994.
59. Rehg JM, Kanade T. Model-based tracking of self-occluding articulated objects. Proceedings of the Fifth IEEE International Conference on Computer Vision ICCV'95; 20–23 June 1995; Boston, MA. New York: IEEE; 1995.
60. Reiman EM, Lane RD, Ahern GL, et al. Neuroanatomical correlates of externally and internally generated human emotion. Am J Psychiatry 1997; 154:918–25.
61. Rohr K. Towards model-based recognition of human movements in image sequences. Computer Vision, Graphics, and Image Processing: Image Understanding 1994; 59:94–115.
62. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 1966; 80:609.
63. Samaras D, Metaxas D, Fua P, Leclerc YG. Variable albedo surface reconstruction from stereo and shape from shading. Proceedings of the IEEE Computer Society on Computer Vision and

Pattern Recognition CVPR'00; 13–15 June 2000; Hilton Head, SC. Washington, DC: IEEE Computer Society; 2000.
64. Scheier MF, Carver CS. Optimism, coping, and health: assessment and implications of generalized outcome expectancies. Health Psychol 1985; 4:219–47.
65. Schmidt K, Cohn JF, Tian YL. Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. Biological Psychology; 2003;65:49–66.
66. Simpson S. Staying sane in space. Sci Am 2000; 282:61–2.
67. Sminchisescu C, Metaxas D, Dickinson S. Improving the scope of deformable model shape and motion estimation. Proceedings of the IEEE Computer Society on Computer Vision and Pattern Recognition CVPR'01; 10–14 Dec 2001; Lihue, HI. Washington, DC: IEEE Computer Society; 2001.
68. Spielberg CD, Gorsuch RL, Lushene RE. Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press; 1970.
69. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935; 18:643–62.
70. Sutton SK, Davidson RJ, Donzella B, et al. Manipulating affective state using extended picture presentations. Psychophysiology 1997; 34:217–26.
71. Taylor GJ, Ryan D, Bagby RM. Toward the development of a new self-report alexithymia scale. Psychother Psychosom 1985; 44: 191–9.
72. Thorne DR, Genser SG, Sing GC, Hegge FW. The Walter Reed performance assessment battery. Neurobehav Toxicol Teratol 1985; 7:415–8.
73. Turk M, Pentland AP. Eigenfaces for recognition. J Cogn Neurosci 1991; 3:71–86.
74. Vogler C, Metaxas D. ASL recognition based on a coupling between HMMs and 3D motion analysis. Proceedings of the IEEE International Conference on Computer Vision, ICCV'98; 4–7 Jan 1998; Bombay, India. New York: IEEE; 1998.
75. Vogler C, Metaxas D. Parallel hidden Markov models for American sign language recognition. Proceedings of the IEEE International Conference on Computer Vision, ICCV'99; 20–25 Sept 1999; Kerkyra, Greece. New York: IEEE; 1999.
76. Vogler C, Metaxas D. Toward scalability in ASL recognition: breaking down signs into phonemes. Proceedings of the International Gesture Workshop, GW'99; 17–19 Mar 1999; Gif-sur-Yvette, France. Paris, France: Springer-Verlag; 1999.
77. Vogler C, Metaxas D. Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods. Proceedings of the IEEE Conference on Systems, Man and Cybernetics, 12–15 October 1997; Orlando, FL. New York: IEEE; 1997:156–61.
78. Yacoob Y, Davis LS. Recognizing human facial expressions from long image sequences using optical flow. IEEE Transactions on Pattern Analysis and Machine Intelligence 1996; 18:638–42.