

RESEARCH

Open Access

Thermal spatio-temporal data for stress recognition

Nandita Sharma^{1*}, Abhinav Dhall¹, Tom Gedeon¹ and Roland Goecke^{1,2}

Abstract

Stress is a serious concern facing our world today, motivating the development of a better objective understanding through the use of non-intrusive means for stress recognition by reducing restrictions to natural human behavior. As an initial step in computer vision-based stress detection, this paper proposes a temporal thermal spectrum (TS) and visible spectrum (VS) video database *ANUStressDB* - a major contribution to stress research. The database contains videos of 35 subjects watching *stressed* and *not-stressed* film clips validated by the subjects. We present the experiment and the process conducted to acquire videos of subjects' faces while they watched the films for the *ANUStressDB*. Further, a baseline model based on computing local binary patterns on three orthogonal planes (LBP-TOP) descriptor on VS and TS videos for stress detection is presented. A LBP-TOP-inspired descriptor was used to capture dynamic thermal patterns in histograms (HDTP) which exploited spatio-temporal characteristics in TS videos. Support vector machines were used for our stress detection model. A genetic algorithm was used to select salient facial block divisions for stress classification and to determine whether certain regions of the face of subjects showed better stress patterns. Results showed that a fusion of facial patterns from VS and TS videos produced statistically significantly better stress recognition rates than patterns from VS or TS videos used in isolation. Moreover, the genetic algorithm selection method led to statistically significantly better stress detection rates than classifiers that used all the facial block divisions. In addition, the best stress recognition rate was obtained from HDTP features fused with LBP-TOP features for TS and VS videos using a hybrid of a genetic algorithm and a support vector machine stress detection model. The model produced an accuracy of 86%.

Keywords: Stress classification; Temporal stress; Thermal imaging; Support vector machines; Genetic algorithms; Watching films

1 Introduction

Stress is a part of everyday life, and it has been widely accepted that stress, which leads to less favorable states (such as anxiety, fear, or anger), is a growing concern to a person's health and well-being, functioning, social interaction, and financial aspects. The term *stress* was coined by Hans Selye, which he defined as 'the non-specific response of the body to any demand for change' [1]. Stress is a natural alarm, resistance, and exhaustion system [2] for the body to prepare for a fight or flight response to either defend or make the body adjust to threats and changes. The body shows stress through symptoms such as frustration, anger, agitation, preoccupation, fear, anxiety, and tenseness

[3]. When chronic and left untreated, stress can lead to incurable illnesses (e.g., cardiovascular diseases [4], diabetes [5], and cancer [6]), relationship deterioration [7,8], and high economic costs, especially in developed countries [9,10]. It is important to recognize stress early to diminish the risks. Stress research is beneficial to our society with a range of benefits, motivating interest and posing technical challenges in computer science in general and affective computing in particular.

Various computational techniques have been used to objectively recognize stress using models based on techniques such as Bayesian networks [11], decision trees [12], support vector machines [13], and artificial neural networks [14]. These techniques have used a range of physiological (e.g., heart activity [15,16], brain activity [17,18], galvanic skin response [19], and skin temperature [12,20]) and physical (e.g., eye gaze [11], facial information [21]) measures

* Correspondence: nandita.sharma@anu.edu.au

¹Information and Human Centred Computing Research Group, Research School of Computer Science, Australian National University, Canberra, ACT 0200, Australia

Full list of author information is available at the end of the article

for stress as inputs. Physiological signal acquisition requires sensors to be in contact with a person, and this can be obtrusive [3]. In addition, the physiological sensors are usually required to be placed on specific locations of the body, and sensor calibration time is usually required as well, e.g., approximately 5 min is needed for the isotonic gel to settle before galvanic skin response readings can be taken satisfactorily using the BIOPAC System [22]. The trend in this area of research is leading towards obtaining symptom of stress measures through less or non-intrusive methods. This paper proposes a stress recognition method using facial imaging and does not require body contact with sensors unlike the usual physiological sensors.

A relatively new area of research is recognition of stress using facial data in the thermal (TS) and visible (VS) spectrums. Blood flow through superficial blood vessels, which are situated under the skin and above the bone and muscle layer of the human body, allows TS images to be captured. It has been reported in the literature that stress can be successfully detected from thermal imaging [23] due to changes in skin temperature under stress. In addition, facial expressions have been analyzed [24] and classified [25-27] using TS imaging. Commonly, VS imaging has been used for modeling facial expressions, and associated robust facial recognition techniques have been developed [28-30]. However, from our understanding, the literature has not developed computational models for stress recognition using both TS and VS imaging together as yet. This paper addresses the gap and presents a robust method to use information from temporal and texture characteristics of facial regions for stress recognition.

Automatic facial expression analysis is a long researched problem. Techniques have been developed for analyzing the temporal dynamics of facial muscle movements. A detailed survey of facial expression recognition methods can be found in [31]. Further, vision-based facial dynamics have been used for affective computing tasks such as pain monitoring [32] and depression analysis [30]. This motivated us to explore vision-based stress analysis where inspiration can be taken from the vast field of facial expression analysis. Descriptors such as the local binary pattern (LBP) have been developed for texture analysis and have been successfully applied to facial expression analysis [25,33,34], depression analysis [30], and face recognition [35]. A particular LBP extension for analysis of temporal data - local binary patterns on three orthogonal planes (LBP-TOP) - has gained attention and is suitable for the work in this study. LBP-TOP provides features that incorporate appearance and motion, and is robust to illumination variations and image transformations [25]. This paper presents an application of LBP-TOP to TS and VS videos.

Various facial dynamics databases have been proposed in the literature. For facial expression analysis, one of the most popular databases is the Cohn-Kanade + [32], which contains facial action coding system (FACS) and generic expression labels. Subjects were asked to pose and display various expressions. There are other databases in the literature which are spontaneous or close to spontaneous, such as RU-FACS [36], Belfast [37], VAM [38], and AFEW [39]. However, these are limited to emotion-related labels which do not serve the problem in the paper, i.e., stress classification. Lucey et al. [32] proposed the UNBC McMasters database comprising video clips where patients were asked to move the arm up and their reaction was recorded. For creating ANUStressDB, subjects were shown stressful and non-stressful video clips. This database is similar to that in [32].

There are various forms of *stressors*, i.e., demands or stimuli that cause stress [23,40-42] validated by self-reports (e.g., self-assessment [43,44]) and observer reports (e.g., human behavior coder [42]). Some examples of stressors are playing video (action) games [45,46], solving difficult mathematical/logical problems [47], and listening to energetic music [45]. Among these stressors are films, which were used to stimulate stress in this work. In this work, we develop a *computed stress measure* [3] using facial imaging in VS and TS. Our work analyzes dynamic facial expressions that are as natural as possible elicited by a typical stressful, tense, or fearful environment from film clips. Unlike the previous work in the literature that uses posed facial expressions for classification [48], the work presented in this paper provides an investigation of spontaneous facial expressions as responses or reactions to environments portrayed by the films.

This paper describes a method for collecting and computationally analyzing data for stress recognition from TS and VS videos. A stress database (ANUStressDB) of videos of faces is presented. An experiment was conducted to collect the data where experiment participants watched stressful and non-stressful film clips. ANUStressDB contains videos of 35 subjects watching film clips that created stressed and not-stressed environments validated by the person. Facial expressions in the videos were stimulated by the film clips. Spatio-temporal features were extracted from the TS and VS videos, and these features were provided as inputs to a support vector machine (SVM) classifier to recognize stress patterns. A hybrid of a genetic algorithm (GA) and SVM was used to select salient divisions of facial block regions and determine whether using the block regions improved the stress recognition rate. The paper compares the quality of the stress classifications produced from using LBP-TOP and HDTP (our thermal spatio-temporal descriptor) features from TS and VS data with and without using facial block selection.

The organization of the paper is as follows: Section 2 presents the experiment for TS, VS, and self-reported data collection. Section 3 describes the facial imaging processing steps for the TS and VS data. The new thermal spatio-temporal descriptor, HDTP, is proposed in Section 4. Stress classification models are described in Section 5. Section 6 presents the results, an analysis of the results, and suggestions for future work.

2 Data collection from the film experiment

After receiving approval from the Australian National University Human Research Ethics Committee, an experiment was conducted to collect TS and VS videos of faces of individuals while they watched films. Thirty-five graduate students consisting of 22 males and 13 females between the ages of 23 and 39 years old volunteered to be experiment participants. Each participant had to understand the experiment requirements from written experiment instructions with the guidance of an experiment instructor before they filled in the consent form. The participant was provided information about the experiment and its purpose from a script to ensure that there was consistency in the experiment information provided across all participants. After providing consent, the participant was seated in front of a LCD display (placed between two speakers). The distance between the screen and subject was in the range between 70 and 90 cm. The instructor started the films, which triggered a blank screen with a countdown of the numbers 3, 2, and 1 transitioning in and out slowly with one before the other. The reason for the countdown display and the blank screen was for participants to move away from their thoughts at the time and get ready to pay attention to the films that were about to start. This approach was like that used in experiments for similar work in [49]. Subsequent to the countdown display, a blank screen was shown for 15 s, which was followed by a sequence of film clips with 5-s blank screens in between. After watching the films, the participant was asked to do a survey, which related to the films they watched and provided validation for the film labels. The experiment took approximately 45 min for each participant. An outline of the process of the experiment for an experiment participant is shown in Figure 1.

Participants watched two types of films either labeled as *stressed* or *not-stressed*. Stressed films had stressful content (e.g., suspense with jumpy music), whereas

not-stressed films created illusions of meditative environments (e.g., swans and ducks paddling in a lake) and had content that was not stressful or at least was relatively less stressful compared with films labeled as stressed. There were six film clips for each type of film. The survey done by experiment participants validated the film labels. The survey asked participants to rate the films they watched in terms of levels of stress portrayed by the film and the degree of tension and relaxation they felt. Participants found the films that were labeled *stressed* as stressful and films labeled *not-stressed* as not stressful with a statistical significance of $p < 0.001$ according to the Wilcoxon test.

While the participants watched the film clips, TS and VS videos of their faces were recorded. A schematic diagram of the experiment setup is shown in Figure 2. TS videos were captured using a FLIR infrared camera (model number SC620, FLIR Systems, Inc. Notting Hill, Australia), and VS videos were recorded using a Microsoft webcam (Microsoft Corporation, Redmond, WA, USA). Both the videos were recorded with a sampling rate of 30 Hz, and the frame width and height were 640 and 480 pixels, respectively. Each participant had a TS and VS video for each film they watched. As a consequence, a participant had 12 video clips made up of six stressed videos and six not-stressed videos. We name the database that has the collected labeled video data and its protocols as the *ANU Stress database (ANUStressDB)*.

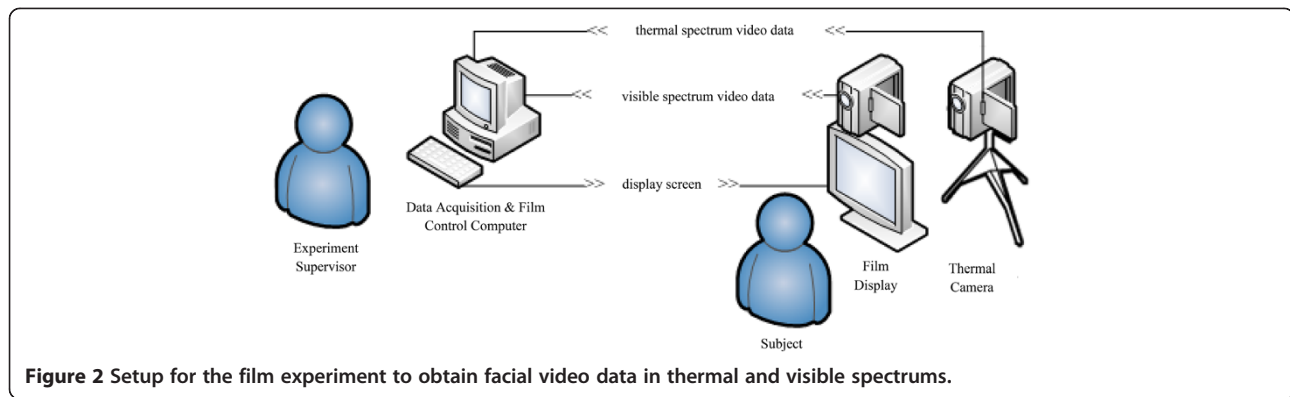
Note the usage of the terms *film* and *video* in this paper. We use the term *film* to refer to a video portraying entertaining content, colloquially called a 'film' or 'movie', which a participant watched during the experiment. We use the term *video* to refer to a visual recording of a participant's face and its movement during the time period while they watched a film. Thus in this paper, a film is something which is watched, while a video is something recorded about the watcher.

3 Face pre-processing pipeline

Facial regions in VS videos were detected using the Viola-Jones face detector. However, facial regions could not be recognized satisfactorily using the Viola-Jones algorithm in thermal spectrum (TS) videos, so a face detection method based on eye coordinates [50,51] and a template matching algorithm was used. A template of a facial region was developed from the first frame of a



Figure 1 An outline of the process followed by each experiment participant in the film experiment.



TS video. The facial region was extracted using the Pretty Helpful Development Functions toolbox for Face Recognition [50-52], which calculated the intraocular displacement to detect a facial region in an image. This facial region formed a template for facial regions in each video frame of the TS videos, which were extracted using MATLAB's Template Matcher system [53]. The Template Matcher was set to search the minimum difference pixel by pixel to find the area of the frame that best matched the template. Examples of facial regions that were detected in the VS and TS videos for a participant are presented in Figure 3.

Facial regions were extracted from each frame of a VS video and its corresponding TS video. Grouped and arranged in order of time of appearance in a video, the facial regions formed *volumes* of the facial region frames. Examples of facial blocks in TS and VS are shown in Figure 4.

4 Spatio-temporal features

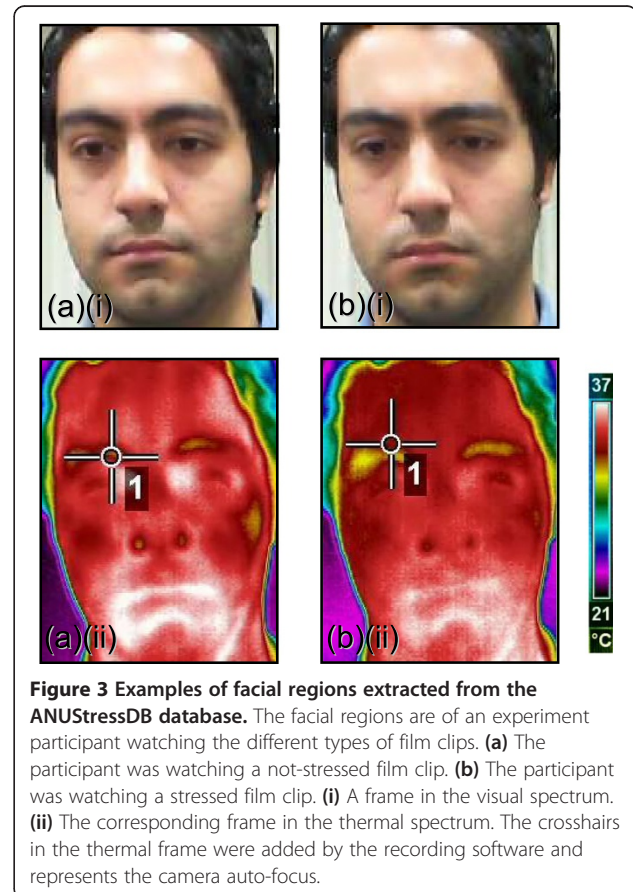
There are claims in the literature that features from segmented image blocks of a facial image region can provide more information than features directly extracted from an image of a full facial region in VS [25]. Examples of full facial regions are shown in Figure 4, and blocks of a full facial region are presented in Figure 5. To illustrate the claim, features from each of the blocks used in conjunction with features from the other blocks in Figure 5 (i) can offer more information than features obtained from Figure 4a (i). The claim aligns with the results from classifying stress based on facial thermal characteristics [23]. As a consequence, the facial regions in this work were segmented into a grid of 3×3 blocks for each video segment, or facial volume, forming 3×3 blocks. A *block* has X , Y , and T components where X , Y , and T represent the width, height, and time components of an image sequence, respectively. Each block represented a division of a full facial block region or facial volume. LBP-TOP features were calculated for each block.

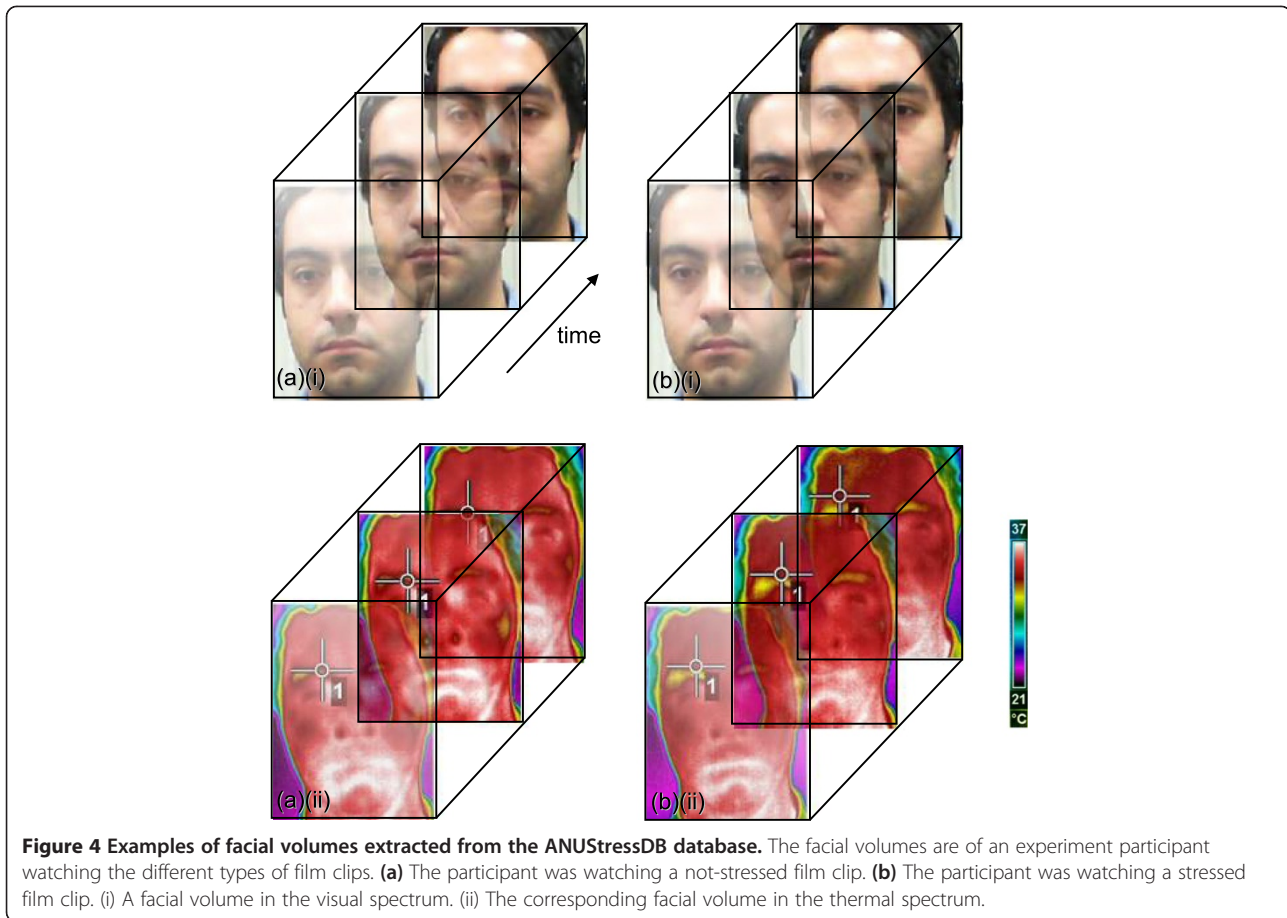
LBP-TOP is the temporal variant of local binary patterns (LBP). In LBP-TOP, LBP is applied to three planes - XY , XT ,

and YT - to describe the appearance of an image, horizontal motion, and vertical motion, respectively. For a center pixel O_p of an orthogonal plane O and its neighboring pixels N_p , a decimal value is assigned to it:

$$d = \sum_O^{XY,XT,YT} \sum_p \sum_{i=1}^k 2^{i-1} I(O_p, N_i) \quad (1)$$

According to a study that investigated facial expression recognition using LBP-TOP features, VS and near-infrared images produced similar facial expression recognition





rates, provided that VS images had strong illumination [33]. Due to the fact that TS videos are defined by colors and different color variations, LBP-TOP features may not be able to fully exploit thermal information provided in TS videos and in particular capture thermal patterns for stress. In addition, LBP-TOP features have been mainly extracted from image sequences of people told to show some facial expression, which is not like the image sequences obtained from our film experiment. In our film experiment, participants watched films and involuntary facial expressions were captured. The recordings may have more subtle facial expressions of the kind of facial expressions analyzed in the literature using LBP-TOP. With the subtleness in facial movement, it is possible that LBP-TOP may not be able to offer as much information for stress analysis. These points motivate the development of a new set of features that exploits thermal patterns in TS videos for stress recognition. We propose a new type of feature for TS videos that captures dynamic thermal patterns in histograms (HDTP). This feature makes use of thermal data in each frame of a TS video of a face over the course of the video.

4.1 Histogram of dynamic thermal patterns

HDTP captures normalized dynamic thermal patterns, which enables individual-independent stress analysis. Some people may be more tolerant to some stressors than others [54,55]. This could mean that some people may show higher degree responses to stress than others. Additionally in general, the baseline for human response can vary from person to person. To consider these characteristics in features used for individual-independent stress analysis, ways have been developed to normalize data for each participant for their type of data [42]. HDTP is defined in terms of a participant's overall thermal state to minimize individual bias in stress analysis.

A HDTP feature is calculated for each facial block region. Firstly, a statistic (consider the standard deviation) is calculated for each facial region frame for a participant for a particular block (e.g., facial block region situated at the top right corner of the facial region in the XY plane) for all the videos. The statistic values from all these frames are partitioned to define empty *bins*. A bin has a continuous value range with a location defined from the statistic values. The bins are used to partition statistic values for each facial block region where the value for

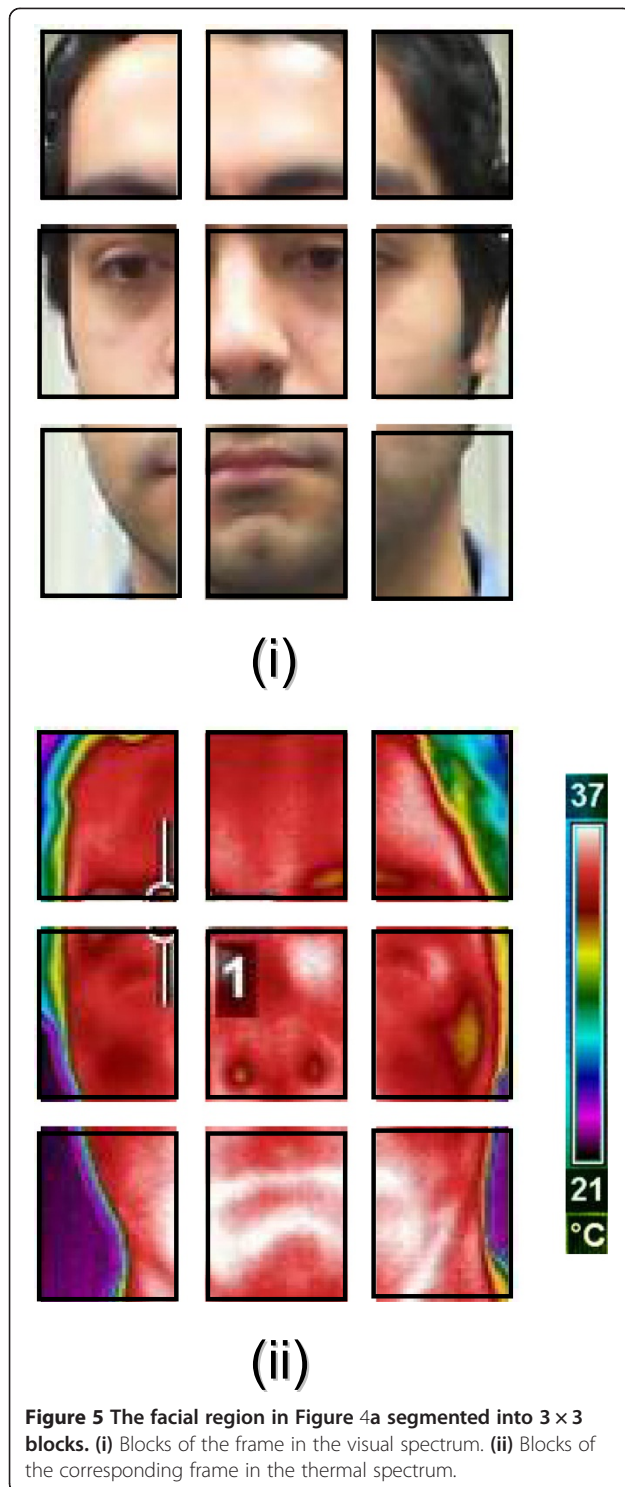


Figure 5 The facial region in Figure 4a segmented into 3 × 3 blocks. (i) Blocks of the frame in the visual spectrum. (ii) Blocks of the corresponding frame in the thermal spectrum.

each bin is the frequency of statistic values in the block that falls within the bounds of the bin range. Consequently, a histogram for each block can be formed from the frequencies. An algorithm presenting the approach for developing histograms of dynamic thermal patterns

in thermal videos for a participant who has a set of facial videos is provided in Figure 6.

As an illustration, consider that the statistic used is the standard deviation and the facial block region for which we want to develop a histogram is situated at the top right corner of the facial region in the XY plane (FBR_1) for video V_1 when a participant P_i was watching film F_1 . In order to create a histogram, the bin locations and sizes need to be calculated. To do this, the standard deviation needs to be calculated for all frames in FBR_1 in all videos (V_{1-12}) for P_i . This will give standard deviation values from which the global minimum and maximum can be obtained and used to calculate the bin location and sizes. Then, the histogram for FBR_1 , for V_1 , and for P_i is calculated by filling the bins with the standard deviation values for each frame in FBR_1 . This method then provides normalized features that also take into account the image and motion, and can be used as inputs to a classifier.

5 Stress classification system using a hybrid of a support vector machine and a genetic algorithm

SVMs have been widely used in the literature to model classification problems including facial expression recognition [27,33,34]. Provided a set of training samples, a SVM transforms the data samples using a nonlinear mapping to a higher dimension with the aim to determine a *hyperplane* that partitions data by class or labels. A hyperplane is chosen based on *support vectors*, which are training data samples that define maximum *margins* from the support vectors to the hyperplane to form the best decision boundary.

It has been reported in the literature that thermal patterns for certain regions of a face provide more information for stress than other regions [23]. The performance of the stress classifier can degrade if irrelevant features are provided as inputs. As a consequence and due to its benefits noted in literature, the classification system was extended to include a feature selection component, which used a GA to select facial block regions appropriate for the stress classification. GAs are inspired by biological evolution and the concept of survival of the fittest. A GA is a global search technique and has been shown to be useful for optimization problems and problems concerning optimal feature selection for classification [56].

The GA evolves a population of candidate solutions, represented by *chromosomes*, using *crossover*, *mutation*, and *selection* operations in search for a better quality population based on some fitness measure. Crossover and mutation operations are applied to chromosomes to achieve diversity in the population and reduce the risk of the search being stuck with a local optimal population. After each generation during the search, the GA selects chromosomes, probabilistically mostly made up of better quality chromosomes, for

Algorithm: HDTP. Find histograms that capture normalized dynamic thermal patterns in thermal videos for a participant

Inputs:

- ♦ p , a participant video data set that represents a set of all videos watched by the participant
- ♦ F , a function that calculates a statistic for a two dimensional matrix

Output:

- ♦ H , a set of histograms for p based on values obtained by F

Method:

```
for each video  $v(p)$ 
  for each block location  $loc(p)$ 
    for each facial block region  $b(v,loc)$ 
      for each frame  $f(b)$ 
         $s_{vbf} \leftarrow F(f(b))$ 
      end for
    end for
  end for
end for
for each block location  $loc(p)$ 
  for each facial block region  $b(loc)$ 
     $bin\_width(b) \leftarrow calculate\_bin\_width(b)$ 
     $bin\_locs(b) \leftarrow calculate\_bin\_locations(bin\_width(b))$ 
  end for
end for
for each video  $v(p)$ 
  for each block location  $loc(p)$ 
    for each facial block region  $b(v,loc)$ 
       $h \leftarrow partition\_data(b, bin\_locs(b))$ 
    end for
  end for
end for
```

Figure 6 The HDTP algorithm captures dynamic thermal patterns in histograms from thermal image sequences.

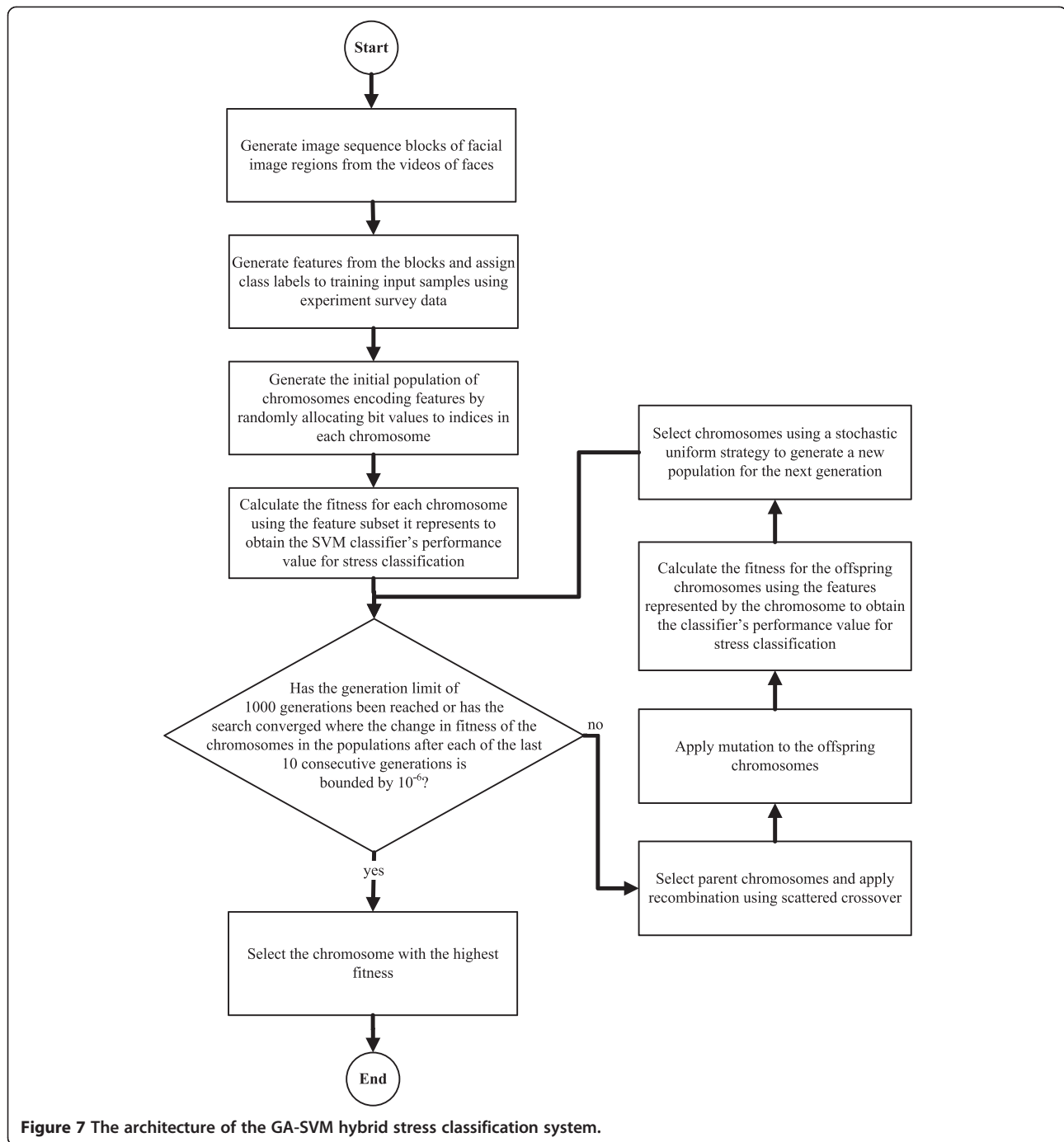
the population in the next generation to direct the search to more favorable chromosomes.

Given a population of subsets of facial block regions with corresponding features, a GA was defined to evolve sets of blocks by applying crossover and mutation operations, and selecting block sets during each iteration of the search to determine sets of blocks that produce better quality SVM classifications. Each block set was represented by a binary fixed-length chromosome where an index or locus symbolized a facial block region; its value or allele depicted whether or not the block was used in the classification and the length of the chromosome matched the number of blocks for a video. The search space had 3×3 blocks (as shown in Figure 5) with

an addition of blocks that overlapped each other by 50%. The architecture for the GA-SVM classification system is shown in Figure 7. The characteristics of the GA implemented for facial block region selection is provided in Table 1.

In summary, various stress classification systems using a SVM were developed which differed in terms of the following input characteristics:

- $VS_{LBP-TOP}$: LBP-TOP features for VS videos
- $TS_{LBP-TOP}$: LBP-TOP features for TS videos
- TS_{HDTP} : HDTP features (as described in Section 4.1) for TS videos
- $VS_{LBP-TOP} + TS_{LBP-TOP}$: $VS_{LBP-TOP}$ and $TS_{LBP-TOP}$



- $VS_{LBP-TOP} + TS_{HDTP}$: $VS_{LBP-TOP}$ and TS_{HDTP}
- $TS_{LBP-TOP} + TS_{HDTP}$: $TS_{LBP-TOP}$ and TS_{HDTP}
- $VS_{LBP-TOP} + TS_{LBP-TOP} + TS_{HDTP}$

These inputs were also provided as inputs to the GA-SVM classification systems to determine whether the system produced better stress recognition rates.

6 Results and discussion

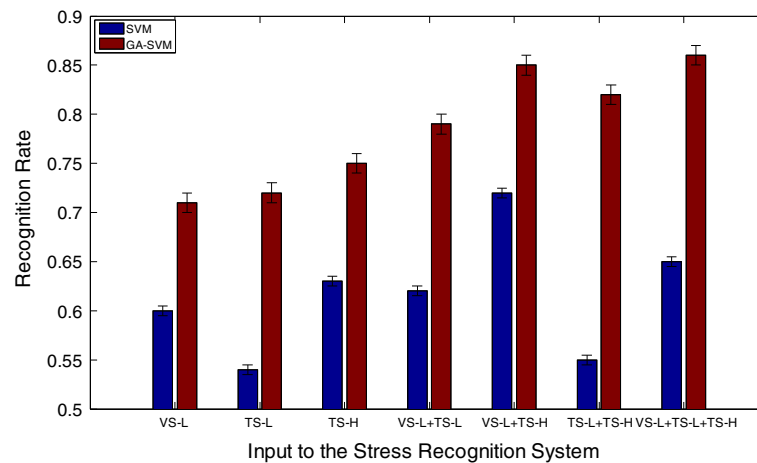
Each of the different features is derived from VS and TS facial videos using LBP-TOP and HDTP facial descriptors on standardized data and provided as inputs to a SVM for stress classification. Facial videos of participants watching stressed films were assigned to the stressed class, and videos associated with not-stressed films were assigned to the

Table 1 GA implementation settings for facial block region selection

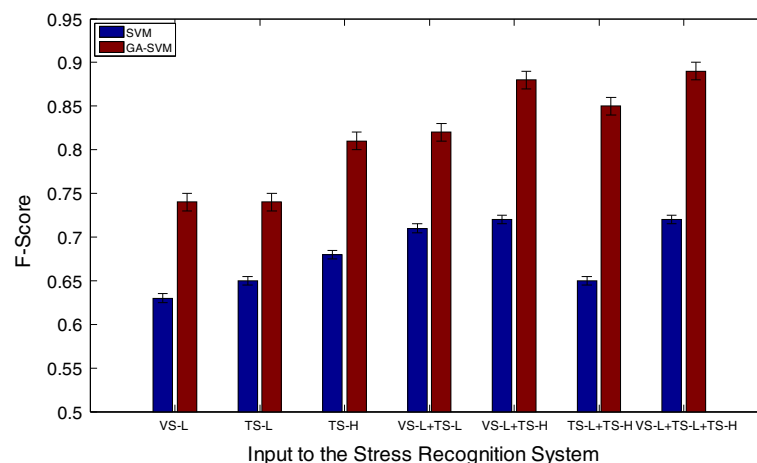
GA parameter	Value/setting
Population size	100
Number of generations	2,000
Crossover rate	0.8
Mutation rate	0.01
Crossover type	Scattered crossover
Mutation type	Uniform mutation
Selection type	Stochastic uniform selection

not-stressed class. Furthermore, their corresponding features were assigned to corresponding classes. Recognition rates and F -scores for the classifications were obtained using 10-fold cross-validation for each type of input. The results are shown in Figure 8.

Results show that when HDTP features for TS videos (TS_{HDTP}) were provided as input to the SVM classifier, there were improvements in the stress recognition measures. The best recognition measures for the SVM were obtained when $VS_{LBP-TOP} + TS_{HDTP}$ was provided as input. It produced a recognition rate that was at least 0.10 greater than the recognition rate for inputs without TS_{HDTP} where the range for recognition rates was 0.13. This provides evidence that TS_{HDTP} had a significant contribution towards the better classification performance and suggests that TS_{HDTP} captured more patterns associated with stress than $VS_{LBP-TOP}$ and $TS_{LBP-TOP}$.



(a)



(b)

Figure 8 Performance measures for SVM and GA-SVM stress recognition systems. The measures were obtained for various input features based on 10-fold cross-validation. The labels on the horizontal axes are shortened to improve readability. L and H stand for LBP-TOP and HDTP, respectively. **(a)** Recognition rate measure for the stress recognition systems. **(b)** F -score measure for the stress recognition systems.

The performance for the classification was the lowest when $TS_{LBP-TOP}$ was provided as input.

The features were also provided as inputs to a GA which selected facial block regions with a goal to disregard irrelevant facial block regions for stress recognition and to improve the SVM-based recognition measures. Performances of the classifications using 10-fold cross-validation on the different inputs are provided in Figure 8. For all types of inputs, GA-SVM produced significantly better stress recognition measures. According to the Wilcoxon non-parametric statistical test, the statistical significance was $p < 0.01$. Similar to the trend observed for stress recognition measures produced by the SVM, TS_{HDTP} also contributed to the improved results in GA-SVM. The best recognition measures were obtained when $VS_{LBP-TOP} + TS_{LBP-TOP} + TS_{HDTP}$ was provided as input to the GA-SVM classifier. The performance of the classifier was highly similar when it received $VS_{LBP-TOP} + TS_{HDTP}$ as inputs with a difference of 0.01 in the recognition rate. Results show that when a combination of at least two of $VS_{LBP-TOP}$, $TS_{LBP-TOP}$, and TS_{HDTP} was provided as input, then it performed better than when only one of $VS_{LBP-TOP}$, $TS_{LBP-TOP}$, or TS_{HDTP} was used.

Further, stress recognition systems provided with TS_{HDTP} as input produced significantly better stress recognition measures than inputs with TS_{HDTP} replaced by $TS_{LBP-TOP}$ ($p < 0.01$). This suggests that stress patterns were better captured by TS_{HDTP} features than $TS_{LBP-TOP}$ features.

In addition, blocks selected by the GA in the GA-SVM classifier for the different inputs were recorded. When $VS_{LBP-TOP}$ was given as inputs to a GA, the blocks that produced better recognition results were the blocks that corresponded to the cheeks and mouth regions on the XY plane. For $VS_{LBP-TOP}$, fewer blocks were selected and they were situated around the nose. On the other hand for TS_{HDTP} , more blocks were used in the classification - nose, mouth, and cheek regions and regions on the forehead were selected by the GA. Future work could extend the investigation by more complex block definitions to find and use more precise regions showing symptoms of stress for classification.

Future work could also investigate other block selection methods different from the GA used in this work. The GA search took approximately 5 min to reach convergence, but it could take longer if the chromosome is extended to encode more general information for a block, e.g., coordinate values and the size for the block. The literature has claimed that a GA usually takes longer execution times than other types of feature selection techniques, such as correlation analysis [57]. Therefore in future, other block selection methods could be investigated that do not require execution times as long as a GA and still produce stress recognition measures comparable to the GA hybrid.

7 Conclusions

The ANU Stress database (ANUStressDB) was presented which has videos of faces in temporal thermal (TS) and visible (VS) spectrums for stress recognition. A computational classification model of stress using spatial and temporal characteristics of facial regions in the ANUStressDB was successfully developed. In the process, a new method for capturing patterns in thermal videos was defined - HDTP. The approach was defined so that it reduced individual bias in the computational models and enhanced participant-independent recognition of symptoms of stress. For computing the baseline for stress classification, a SVM was used. Facial block regions selected informed by a genetic algorithm improved the rates of the classifications regardless of the type of video - videos in TS or VS. The best recognition rates, however, were obtained when features from TS and VS videos were provided as inputs to the GA-SVM classifier. In addition, stress recognition rates were significantly better for classifiers provided with HDTP features instead of LBP-TOP features for TS. Future work could extend the investigation by developing features for facial block regions to capture more complex patterns and examining different forms of facial block regions for stress recognition.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Information and Human Centred Computing Research Group, Research School of Computer Science, Australian National University, Canberra, ACT 0200, Australia. ²Vision & Sensing Group, Information Sciences & Engineering, University of Canberra, Bruce ACT 2601, Australia.

Received: 2 December 2012 Accepted: 13 May 2014

Published: 4 June 2014

References

1. H Selye, The stress syndrome. *Am. J. Nurs.* **65**, 97–99 (1965)
2. L Hoffman-Goetz, BK Pedersen, Exercise and the immune system: a model of the stress response? *Immunol. Today* **15**, 382–387 (1994)
3. N Sharma, T Gedeon, Objective measures, sensors and computational techniques for stress recognition and classification: a survey. *Comput. Methods Prog. Biomed.* **108**, 1287–1301 (2012)
4. GE Miller, S Cohen, AK Ritchey, Chronic psychological stress and the regulation of pro-inflammatory cytokines: a glucocorticoid-resistance model. *Health Psychology Hillsdale* **21**, 531–541 (2002)
5. RS Surwit, MS Schneider, MN Feinglos, Stress and diabetes mellitus. *Diabetes Care* **15**, 1413–1422 (1992)
6. L Vitetta, B Anton, F Cortizo, A Sali, Mind body medicine: stress and its impact on overall health and longevity. *Ann. N. Y. Acad. Sci.* **1057**, 492–505 (2005)
7. JA Seltzer, D Kalmuss, Socialization and stress explanations for spouse abuse. *Social Forces* **67**, 473–491 (1988)
8. PR Johnson, J Indvik, Stress and violence in the workplace. *Employee Counsell. Today* **8**, 19–24 (1996)
9. The American Institute of Stress. (05/08/10), America's no. 1 health problem - why is there more stress today? <http://www.stress.org/>. Accessed 5 August 2010
10. Lifeline Australia, Stress costs taxpayer \$300K every day, 2009. <http://www.lifeline.org.au> Accessed 10 August 2010
11. W Liao, W Zhang, Z Zhu, Q Ji, A real-time human stress monitoring system using dynamic Bayesian network, in *Computer Vision and Pattern Recognition - Workshops, CVPR Workshops*. San Diego, CA, USA, 25 June 2005.

12. J Zhai, A Barreto, Stress recognition using non-invasive technology, in *Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference FLAIRS*. Melbourne Beach, Florida, 2006, pp. 395–400
13. J Wang, M Korczykowski, H Rao, Y Fan, J Pluta, RC Gur, BS McEwen, JA Detre, Gender difference in neural response to psychological stress. *Soc. Cogn. Affect. Neurosci.* **2**, 227 (2007)
14. N Sharma, T Gedeon, in *Stress Classification for Gender Bias in Reading - Neural Information Processing vol. 7064*, ed. by B-L Lu, L Zhang, J Kwok (Springer, Berlin, 2011), pp. 348–355
15. T Ushiyama, K Mizushige, H Wakabayashi, T Nakatsu, K Ishimura, Y Tsuboi, H Maeta, Y Suzuki, Analysis of heart rate variability as an index of noncardiac surgical stress. *Heart Vessel.* **23**, 53–59 (2008)
16. H Seong, J Lee, T Shin, W Kim, Y Yoon, The analysis of mental stress using time-frequency distribution of heart rate variability signal, in *Annual International Conference of Engineering in Medicine and Biology Society, 2004*. San Francisco, CA, USA, 1–4 September 2004, vol 1, pp. 283–285
17. DA Morilak, G Barrera, DJ Echevarria, AS Garcia, A Hernandez, S Ma, CO Petre, Role of brain norepinephrine in the behavioral response to stress. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **29**, 1214–1224 (2005)
18. M Haak, S Bos, S Panic, LJM Rothkrantz, Detecting stress using eye blinks and brain activity from EEG signals, in *Proceeding of the 1st Driver Car Interaction and Interface (DCII 2008)*. Chez Technical University, Prague, 2008.
19. Y Shi, N Ruiz, R Taib, E Choi, F Chen, Galvanic skin response (GSR) as an index of cognitive load, in *CHI '07 extended abstracts on Human factors in computing systems*. San Jose, CA, USA, 2007, 28 April - 3 May 2007, pp. 2651–2656
20. S Reisman, Measurement of physiological stress, in *Bioengineering Conference*. 1997, 4–6 April 1997, pp. 21–23
21. DF Dinges, RL Rider, J Dorrian, EL McGlinchey, NL Rogers, Z Cizman, SK Goldenstein, C Vogler, S Venkataraman, DN Metaxas, Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviat. Space Environ. Med.* **76**, B172–B182 (2005)
22. BIOPAC Systems Inc, BIOPAC Systems, 2012. <http://www.biopac.com/>. Accessed 10 February 2011
23. P Yuen, K Hong, T Chen, A Tsitiridis, F Kam, J Jackman, D James, M Richardson, L Williams, W. Oxford, Emotional & physical stress detection and classification using thermal imaging technique, in *3rd International Conference on Crime Detection and Prevention (ICDP 2009)*. London, 2009, 3 December 2009, pp. 1–6
24. S Jarlier, D Grandjean, S Delplanque, K N'Diaye, I Cayeux, MI Velazco, D Sander, P Vuilleumier, KR Scherer, Thermal analysis of facial muscles contractions. *IEEE Trans. Affect. Comput.* **2**, 2–9 (2011)
25. G Zhao, M Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 915–928 (2007)
26. B Hernández, G Olague, R Hammoud, L Trujillo, E Romero, Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Comput. Vis. Image Underst.* **106**, 258–269 (2007)
27. L Trujillo, G Olague, R Hammoud, B Hernandez, Automatic feature localization in thermal images for facial expression recognition, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops*. San Diego, CA, USA, 2005, 20, 21 and 25 June 2005, p. 14
28. PK Manglik, U Misra, HB Maringanti, Facial expression recognition, in *IEEE International Conference on Systems, Man and Cybernetics*. 2004, The Hague, Netherlands, 10–13 October 2004, pp. 2220–2224
29. N Neggaz, M Besnassi, A Benyettou, Application of improved AAM and probabilistic neural network to facial expression recognition. *J. Appl. Sci.* **10**, 1572–1579 (2010)
30. G Sandbach, S Zafeiriou, M Pantic, D Rueckert, Recognition of 3D facial expression dynamics. *Image Vis. Comput.* **30**, 762–773 (2012)
31. Z Zeng, M Pantic, GI Roisman, TS Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 39–58 (2009)
32. P Lucey, JF Cohn, KM Prkachin, PE Solomon, I Matthews, Painful data: The UNBC-McMaster shoulder pain expression archive database, in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*. 2011, Santa Barbara, CA, USA, 21–25 March 2011, pp. 57–64
33. M Taini, G Zhao, SZ Li, M Pietikainen, Facial expression recognition from near-infrared video sequences, in *19th International Conference on Pattern Recognition (ICPR)*. 2008, Tampa, Florida, USA, 8–11 December 2008, pp. 1–4
34. P Michel, RE Kaliouby, Real time facial expression recognition in video using support vector machines, in *the Proceedings of the 5th International Conference on Multimodal Interfaces, 2003*. Vancouver, British Columbia, Canada, 5–7 November 2003
35. T Ahonen, A Hadid, M Pietikainen, Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 2037–2041 (2006)
36. MS Bartlett, GC Littlewort, MG Frank, C Lainscsek, IR Fasel, JR Movellan, Automatic recognition of facial actions in spontaneous expressions. *J. Multimed.* **1**, 22–35 (2006)
37. E Douglas-Cowie, R Cowie, M Schröder, A new emotion database: considerations, sources and scope, in *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000*, pp. 39–44
38. M Grimm, K Kroschel, S Narayanan, The Vera am Mittag German audio-visual emotional speech database, in *IEEE International Conference on Multimedia and Expo*. Hannover, Germany, 23–26 June 2008, 2008, pp. 865–868
39. A Dhall, R Goecke, S Lucey, T Gedeon, A semi-automatic method for collecting richly labelled large facial expression databases from movies. *IEEE Multimedia* **19**, 34–41 (2012)
40. J Zhai, A Barreto, Stress detection in computer users based on digital signal processing of noninvasive physiological variables, in *Proceedings of the 28th IEEE EMBS Annual International Conference*. 2006, New York City, NY, USA, 30 August - 3 September 2006, pp. 1355–1358
41. N Hjortskov, D Rissén, A Blangsted, N Fallentin, U Lundberg, K Søgaard, The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* **92**, 84–89 (2004)
42. JA Healey, RW Picard, Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transport. Syst.* **6**, 156–166 (2005)
43. A Niculescu, Y Cao, A Nijholt, Manipulating stress and cognitive load in conversational interactions with a multimodal system for crisis management support, in *Development of Multimodal Interfaces: Active Listening and Synchrony* (Springer, Dublin Ireland, 2010), pp. 134–147
44. LM Vizer, L Zhou, A Sears, Automated stress detection using keystroke and linguistic features: an exploratory study. *Int. J. Hum. Comput. Stud.* **67**, 870–886 (2009)
45. T Lin, L John, Quantifying mental relaxation with EEG for use in computer games, in *International Conference on Internet Computing, 2006*. Las Vegas, NV, USA, 26–29 June 2006, pp. 409–415
46. T Lin, M Omata, W Hu, A Imamiya, Do physiological data relate to traditional usability indexes? in *Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future* (Narrabundah, Australia, 2005), pp. 1–10
47. WR Lovallo, *Stress & Health: Biological and Psychological Interactions* (Sage Publications, Inc., California, 2005)
48. P Lucey, JF Cohn, T Kanade, J Saragih, Z Ambadar, I Matthews, The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (San Francisco, CA, USA, 2010), pp. 94–101
49. JJ Gross, RW Levenson, Emotion elicitation using films. *Cognit. Emot.* **9**, 87–108 (1995)
50. V Struc, N Pavesic, The complete Gabor-Fisher classifier for robust face recognition. *EURASIP Advances in Signal Processing* **2010**, 26 (2010)
51. V Struc, N Pavesic, Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica (Vilnius)* **20**, 115–138 (2009)
52. V Struc, *The PhD Toolbox: Pretty Helpful Development Functions for Face Recognition*, 2012. http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/PhDFace/. Accessed 12 September 2012
53. Mathworks, Vision TemplateMatcher System Object R2012a. (2012). <http://www.mathworks.com.au/help/vision/ref/vision.templatematcherclass.html>. Accessed 12 September 2012
54. APA, *American Psychological Association, Stress in America* (APA, Washington, DC, 2012)
55. CJ Holahan, RH Moos, Life stressors, resistance factors, and improved psychological functioning: an extension of the stress resistance paradigm. *J. Pers. Soc. Psychol.* **58**, 909 (1990)

56. H Frohlich, O Chapelle, B Scholkopf, Feature selection for support vector machines by means of genetic algorithm, in *15th IEEE International Conference on Tools with Artificial Intelligence*. Sacramento, California, USA, 3–5 November 2003, 2003, pp. 142–148
57. L Yu, H Liu, Feature selection for high-dimensional data: a fast correlation-based filter solution, in *12th International Conference on Machine Learning*. Los Angeles, CA, 23–24 June 2003, 2003, pp. 856–863

doi:10.1186/1687-5281-2014-28

Cite this article as: Sharma et al.: Thermal spatio-temporal data for stress recognition. *EURASIP Journal on Image and Video Processing* 2014 **2014**:28.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
