AALBORG UNIVERSITY

DENMARK

Thermal Super-Pixels for Bimodal Stress Recognition

Irani, Ramin; Nasrollahi, Kamal; Dhall, Abhinav; Moeslund, Thomas B.; Gedeon, Tom

Link to publication from Aalborg University

Thermal Super-Pixels for Bimodal Stress Recognition

Ramin Irani[1], Kamal Nasrollahi[1], Abhinav Dhall[2], Thomas B. Moeslund[1], and Tom Gedeon[3]

[1] Visual Analysis of People (VAP) Laboratory, Aalborg University, Denmark
e-mail: {ri, kn, tbm}@create.aau.dk

[2] Computational Health & Informatics Lab (CHIL), University of Waterloo, Canada
e-mail: abhinav.dhall@uwaterloo.ca

[3] Human Centred Computing (HCC), Australian National University, Australia
e-mail: tom@cs.anu.edu.au

Abstract—Stress is a response to time pressure or negative environmental conditions. If its stimulus iterates or stays for a long time, it affects health conditions. Thus, stress recognition is an important issue. Traditional systems for this purpose are mostly contact-based, i.e., they require a sensor to be in touch with the body which is not always practical. Contact-free monitoring of the stress by a camera [1], [2] can be an alternative. These systems usually utilize only an RGB or a thermal camera to recognize stress. To the best of our knowledge, the only work on fusion of these two modalities for stress recognition is [3] which uses a feature level fusion of the two modalities. The features in [3] are extracted directly from pixel values. In this paper we show that extracting the features from super-pixels, followed by decision level fusion results in a system outperforming [3]. The experimental results on ANUstressDB database show that our system achieves 89% classification accuracy.

Keywords—Stress Recognition, Facial Expression, RGB Images, Thermal Images, Super-pixels.

I. Introduction

Nowadays stress is a major problem in the human society. Usually the reason for stress is recognized as time pressure. For example, when we want to perform a task within a given period, while we do not have enough time, a set of physiological reactions, like, heartbeat and respiration rates increase, which indicate a stressful situation [4]. This situation, however, may not be the same for different people, as stress is subjective. In other words, stress depends more on changes in specific physiological signs, not on conditions/events themselves [4]. It is also argued that different people may experience different conditions or events and hence their physiological signs may change differently.

Traditional stress recognition systems, which are based on self-report or measure physiological signals using invasive sensors have some limitations. For example: these systems are unable to monitor the subjects instantaneously and continuously [5]. Some of the systems are based on self-reporting or saliva test. Self-reporting based systems at times may not be able to recognize the stress on short duration of time. As a consequence, to overcome these problems researchers nowadays tend to measure stress using contact-less sensors such as RGB and thermal cameras.

Since stress is associated with physical appearance, some researchers utilized the physical symptoms as clues for stress detection. For example, in [6], [7] a model for monitoring the subjects based on deformation of lips, mouth and eyebrows due to the stress has been presented. Liao et al [1] proposed an approach to recognize the stress using some visual features like blinking frequency, average eyes closure speed and percentage of saccadic eye movement. In another work [8] authors detected the stress by tracking 3D facial expressions. In a recent work, Gao et al [9] presented a contact-less real time system for detecting stress in vehicle drivers. It functioned by considering two negative basic facial expressions (anger and disgust) based on the Ekman theory [10] and applied Local discrete cosine transform [11] and scale invariant feature transform [12] as features.

Since physical appearance is not as reliable as physiological response to stressors, many researchers are interested in employing these symptoms for stress detection. Recently, imaging techniques like RGB video recorder or thermal imaging have been employed for contact-less measuring of physiological signals, e.g., Hearthbeat rate [13], [14], respiration rate [15], [16] and muscle fatigue [17] which promise contact-less-based measuring of the stress as well.

Hyperspectral imaging [5], thermal imaging and RGB Imaging have been used to probe the stress using physiological features. Pavlidis et al [18], [19] are the first researchers who measured the stress with a contact-less thermal sensor. The principle of their work is based on the fact that mental stress increases the blood flow in the forehead region. Thus, they applied a contact-less blood flow measurement on 10% hottest pixel of the ROI for monitoring the stress. To quantify the stress level, in another interesting work, Shastri et al. [20] captured thermal imaging based data for measuring the transient perspiration, which is also known as a physiological functions. The drawback of this method is that stress cannot be measured in scenarios, where subjects experience heavy sweating due to hot environment or during doing exercise.

In [2] Frédéric et. al. proposed a system based on instantaneous pulse rate signal extracted from imaging photoplethysmography. The proposed algorithm derives the Heart Rate Variability (HRV) from a webcam and detects the stress by analysing the HRV changes due to stress.

Employed vision based systems for stress recognition usually use only one of the RGB or thermal imaging techniques. To employ the opportunities of fusing the two modalities, a recent literature [3] presented a computational model using

the information from both thermal and RGB imaging. They proposed a new descriptor named Histogram of Dynamic Thermal Patterns (HDTP). However,they could not achieve more than 65% accuracy. Nevertheless, such accuracy has been improved to 85% by combining RGB and thermal imaging features as input of a Genetic Algorithm (GA)-Support Vector Machine (SVM) classifier.

In thermal images an unique color is assigned to the pixels with similar temperature (Figure 1.a). This gives rise to a formation of regions on the facial block. An advantage of thermal based face analysis over RGB is that it is less effeced by noise in the facial parts location detection. Extracting features in the ub-regions (block-wise) instead of the entire face (holistic-level) improves the accuracy [21] in RGB images based face analysis. Generally, a particular block may cover a facial part or two adjacent blocks may contain a particular facial part. However, this is not guaranteed for thermal images as the sub-regions in thermal areas donot strictly adhere to facial part boundaries. Another reason for this can be that the if the thermal image is divided into fixed blocks, it may result into a block containing different (complete/incomplete) thermal regions, which may not have any correlation. Furthermore, due to the process of image capturing a sensor quantizes a natural continous signal (image) into pixels. Motivated by these obeservations, in this paper we propose to represent a thermal image as a group of super-pixels. A super-pixels is a group of adjacent pixels which have similar characteristic and special information (Figure 1.b). Super-pixel representation has been used for face recognition [22]. In the case of thermal images, super-pixels are a group of pixels with similar color (temperature) which seem like a more natural representation for thermal images as compared to dividing images into non-overlapping blocks. This method not only groups the adjacent pixels with high correlation but also increases the speed of processing. Our experimental results show promising outcomes and are in agreement with the state-of-the-art method of [3].
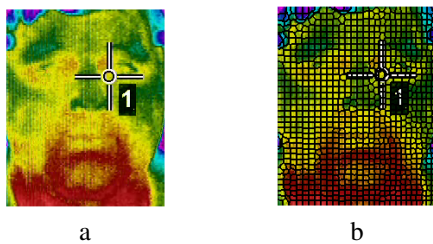

Fig. 1.   A typical facial Region (a) and its corresponding super-pixels (b)

Super-pixels are the results of perceptual grouping of pixels and involve more information and provide better image alignment compared to using a single pixel alone [23]. Mapping from a pixel grid to super-pixel, holds desirable properties, like, computational efficiency, perceptual meaningfulness, over segmentation, and efficient graph representations [24]. Super-pixels share some properties like texture distribution or color

similarity. Specially, this attribute can be helpful in thermal image analyzing, because we are interested in temperature of sub-regions instead of points.

In the recent years, there has been progress in super-pixel creating algorithms [25] [26]. A detailed procs and cons of various super-pixel algorithms is presented in [23]. In this work, for computing super-pixels, Linear Spectral Clustering (LSC) method [27] is followed. The reason for chosing LSC is its ability to produce fast compact and uniform super-pixels.

The rest of the paper is organized as follows: Section II explains the details of the proposed system, Section III discusses the experimental results, and finally, Section IV concludes the paper.

II. THE PROPOSED SYSTEM

The block diagram of the proposed system is shown in Figure 2. The test subjects are filmed by a RGB camera that is synchronized with a thermal camera in parallel. These two types of video streams go through three different steps: 1)Face region detection and quality assessment, 2) Feature extraction, and 3) Classification and fusion .

Since the data collected by the two types of cameras are different in nature, the applied algorithms in the first two steps are different. For RBG images, recognizing stress is similar to [1], that is, first facial region is detected by the Viola Jones (VJ) face detector. Then, the face regions with less correlation are removed using a face quality assessment algorithm. Finally, Local Binary Patterns (LBP) [28] are extracted from the remaining facial regions and are used as feature points. However for detecting the face area in the thermal images, we use a template matcher for face region detection as proposed in [31]. Then, we compute the LSC super-pixel algorithm, instead of directly computing a facial descriptor. Further, the mean values of the generated super-pixels are used as the facial features. Having extracted the facial features from two types of inputs, we use a support vector machine (SVM) classifier for producing classification scores for each type of input. These scores are finally fused at decision level to recognize the stress. These steps are explained in detail in the following subsections.

A. Step 1: face detection and quality assessment

1) RGB Data: The first step of stress recognition in RGB videos is cropping the face region. We used the VJ face detection algorithm [29] for this purpose. In order to decrease the error of the algorithm, if it can not find a face in the current frame, we use the position of the frame in the previous frame as the position of the face in the current frame. Considering the fact that in our employed database (discussed later) the subjects' face does not have considerable head pose changes and movements within short period of time, this method seems working and reducing the error of the face detection algorithm, when it fails to detect faces. Furthermore, if there are more than one region detected as face, we utilize the information
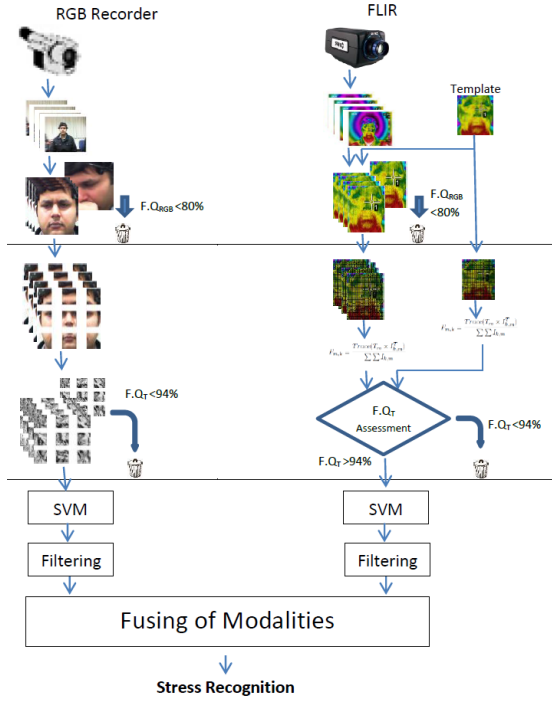
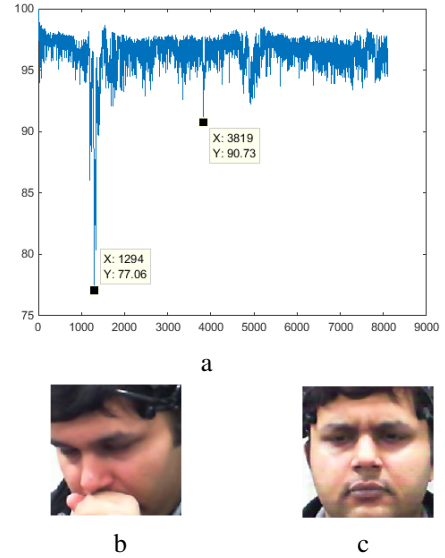Fig. 2.   The block diagram of proposed bimodal system

Fig. 3.   Quality assessment, a. Correlation of the frames with a chosen template for all the frames in subject 1 video sequence, b. a frame with correlation less than 80% c. a frame with correlation larger than 80%

about the setup (discussed later in III) to keep only the one which is closest in size to $v \times w$. The values of v and w are determined experimentally, based on the distance of the subjects from camera.

Finally, we employ a face quality assessment technique for detecting the frames with incorrect face region (Figure 3. To do so, we use the first detected face in the first frame as a reference face and discard all the other faces that are not similar enough to this reference face (less than 80%). The similarity is calculated using the following correlation:

$$S_{RBG} = \frac{\sum_m^M \sum_n^N \left(\mathbf{A}_{mn} - \bar{A}\right)\left(\mathbf{B}_{mn} - \bar{B}\right)}{\sqrt{\sum_m^M \sum_n^N (\mathbf{A}_{mn} - \bar{A})^2 . \sum_m^M \sum_n^N (\mathbf{B}_{mn} - \bar{B})^2}}$$
$$\times 100\%$$
$$(1)$$

in which, $\mathbf{A}$ is the template/reference face, $\bar{A}$ is the average grey level in the reference image, $\mathbf{B}$ is the face in the current frame, $\bar{B}$ is the average grey level of the face in the current frame, and M & N are the number of rows and columns of frames, respectively, (template image size = columns $\times$ rows).

Figure 3.a shows a correlation curve obtained by the above formula for the entire faces of a video sequence and two faces. The first face (Figure 3.b) has been discarded while the second one (Figure 3.c) has been kept.

It should be mentioned that when we discard a frame/face in RGB video sequence, it's corresponding frame in the thermal video sequence should also be discarded. It is an essential

condition to keep the synchronization between the modalities.

2) Thermal Data: Before applying the LSC super-pixel algorithm on the thermal images, face localization in the thermal images is required. Since the VJ algorithm, which were applied on the RGB frames, is not useful in this case, we used a template matcher [30]. The template is created manually for each thermal video sequence. The facial region on one frame (the reference frame) is cropped and then used to find the facial regions in the rest of the frames using the Yue Wu algorithm of in [30] which is based on correlation. Figure 4.b shows the correlation values between the template and the face region in the current frame (Figure 4.a). The brightest point on the correlation map (Figure 4.b) indicates center of the face region that should be cropped. Figure 4.c shows the detected and cropped face region.

B. Step 2: Feature extraction

1) Extracting features from RGB data: It is discussed in [21] that for RGB images it is be better to extract facial features from individual non-overlapping blocks than at the holistic level. In this work, similar to [3] in each frame the facial region is segmented into a grid of 3×3 blocks. Next, LBP features are computed for each block [28]. LBP has been successfully utilized in many facial analysis systems, like [31] [32] Figure 5.a and 5.b show an input image divided into 3 by 3 blocks and their corresponding LBP counterparts.

2) Extracting features from thermal data : To extract features from thermal images, we first apply the super-pixel technique of [27] to segment the face regions to small none-overlapping pieces. This technique divides the facial region to some sub-regions (Figure1.b) that unlike blocked RGB images (Figure5.a), each sub-region includes pixels with mostly
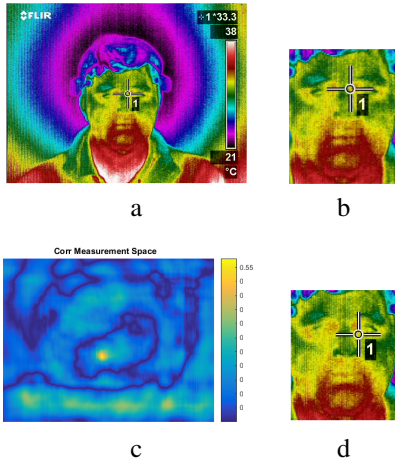
Fig. 4. Template matching process for detecting facial region in thermal images: a. a frame from the ANU StressDB database, b. Template, c. correlation map, d. the detected Facial region
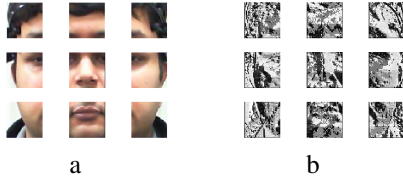

Fig. 5. Facial sub-regions in a. RGB frames, b. Corresponding LBP features

similar color (hence similar temperature in this case). Such property, in addition to the fact that the stress has direct correlation with skins temperature, made it possible to consider the mean temperature of each super-pixel as feature of the corresponding sub-region. The region of each super-pixel is determined using a matrix named Label with the size equal to the size of a face. Label assigns an integer to each super-pixel such that:

$$Label = \bigcup_{k=1}^{K} k.I_k \qquad (2)$$

in which:

$$I_k = \begin{cases} 1, & \text{if } P_{i,j} = k \\ 0, & \text{Otherwise} \end{cases} \qquad (3)$$

The thermal feature F_m for each frame is given by:

$$F_{m,k} = \frac{Tr(T_m \times I_{k,m}^T)}{\sum \sum I_{k,m}} \qquad (4)$$

where, $F_{m,k}$ is k_{th} element of the feature vector of m_{th} frame, T_m is m_{th} thermal frame, $I_{k,m}^T$ is transport matrix of k_{th} super-pixel in frame m, $P_{i,j}$ is a pixel on the thermal image of T_m with coordinate i, j, $\bigcup$ is a union function, and Tr is Trace function.

Similar to RGB modality we apply a quality assessment to thermal modality. The quality assessment is however not applied to the thermal images, but to their features, as these features are one dimensional (average of blocks) and are much easier to process than the images. To do so, we use the correlation scores obtained by Equation 5. The difference between this correlation and the one used for the RGB modality is that we here have replaced gray levels with the mean of super-pixels. In addition, since the applied features (temperature) involves values with small variation, an exponential function and with a factor of α has been considered to depict the frames with less quality with decreasing their corresponding scores more faster than high quality frames, as in:

$$S_T = \exp\left(\alpha \times \left(\frac{\sum_{i=0}^{N-1}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{N-1}(x_i - \bar{x})^2 . \sum_{i=0}^{N-1}(y_i - \bar{y})^2}} - 1\right)\right) \times 100 \qquad (5)$$

where, x is the feature vector of the template frame, $\bar{x}$ is the average of template feature vector, y is the feature vector of source frame, $\bar{y}$ is the average of source features vector, and N is the size of feature vector. The value of S_T varies within 0 and 100, such that larger values of "S_T" represent a strong relationship between the two images. The features (also frames) with score (S_T) less than 94% were removed. Figure 6.a illustrates the score of a video sequence. Figures 6.b and 6.c show corresponding frames of the spots marked on figure 6.a. It can be seen that the frame with a score less than 94% is not correctly matching with the template.
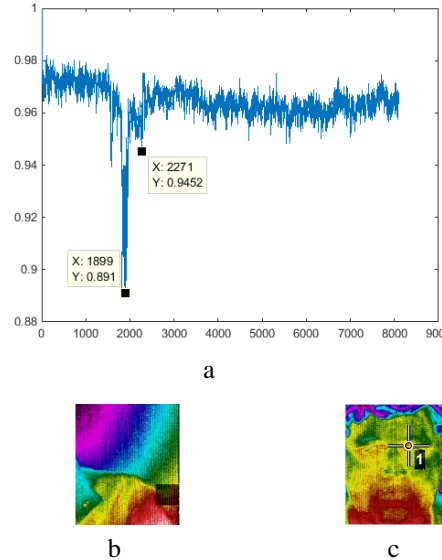

Fig. 6. Quality assessment of thermal images, a. Similarity of a thermal segment for subject 1, b. detected face region with score 89%, c. detected face region with score 95% .

C. Fusing and classification

Motivated by the successful application of SVM in different vision algorithms [33], [34], we have decided to use it for classifying our features. Two separate SVMs have been used for the classification of the features extracted from the different modalities. The output of these SVMs need to be fused to make a decision if the test subjects are in stress in the current RGB and its corresponding thermal frames or not. Since stress is a continues phenomenon and cannot vary abruptly, we have assumed that the level of the stress does not change in short periods (here experimentally obtained period of four seconds has been used). To reflect this temporal period, we have applied a median and then a mean filter to the output of each SVM. The median filter removes the outlier scores of the SVMs, while the mean filter (moving average) aggregates the scores over temporal period of the stress. To apply the moving average, the outputs of median filter are windowed with length of N and an overlapping factor of N-1. Finally, the output of moving average filter is weighted and fused using the following equation:

$$S_{Modal} = tanh(\gamma.(\omega_1.S_{RGB} + \omega_2.S_T + Threshold)) \quad (6)$$

where, S_{RGB} is output of RGB SVM modal, S_T is output of thermal SVM modal, S_{Modal} is the final output after fusion, ω_1 and ω_2 are weight coefficients of RGB and thermal inputs of the fusion, and $Threshold$ is a threshold for making decision if the frame is stressful or not. Frames with corresponding value less than threshold are stressful frame and those larger than threshold are non-stressful.

III. EXPERIMENTAL RESULTS AND DISCUSSION

This proposed system, has been tested on the only database for stress recognition that contains images of both RGB and thermal modalities, ANUStressDB. This database has been collected at Australian National University (ANU) [3] and involves 35 subjects, composed of 22 males and 13 females, between 23 and 39 years old. The thermal and RGB modalities were captured by a FLIR infrared camera and a Microsoft webcam, respectively. Both cameras were working at 30 frame per second at a 640x480 pixels resolution. We set the values of v and w (of section II-A.1) to 110 pixels each.

Instructors played a film with a collection of negative and positive clips as stress stimulator. The clips are separated by displaying 5 seconds blank screen in-between the clips in order to neutralize the participants' emotion (state of mind) before displaying the next movie. Therefor, in the ground truth data, we assigned all the frames as "stressed/unstressed" when the label of the film is "stressed/unstressed".At the end of the experiment, participants were asked to fill a questionnaire survey for the validation of the experiment.

For classifying the extracted features using SVM, 60% of the samples from each modality were selected for training, and the rest for testing. Since the stress is a temporal process, besides considering the SVM scores directly, we have considered applying some temporal post-processing technique which consider a kind of history for the current frame to be involved in the decision making about the current frame to be classified as stress or not. For this purpose, we have simply looked into mean and median filters. In other words, to decide whether or not the current frame is of a stressful situation, besides looking into the SVM score of the current frame, we apply once a mean and once a median filter to the SVM scores of the frames located within a neighborhood of the current frame.

Table 1 shows the results obtained by SVMs for each modality without any post-processing (the second column), with mean and median filters applied to the results of the SVM (third and fourth columns, respectively), and after fusing the post-processed (mean-filtered) results of both modalities (fifth column).

Table 1. Comparing improvement of the results in each step of post processing filters and fusing

Modal	SVM	Median	Moving Average	Fusion
RGB	60%	60%	62%	89%
Thermal	82%	84%	86%	

Figure 7 shows the results of the proposed system aganist the ground thruth after fusing the scores coming from both modalities.
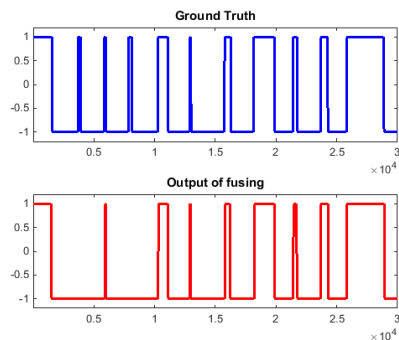

Fig. 7. Comparing the ground truth with final result after fusing the modalalities

Table 2 shows the results of comparison of the modality fusion of the proposed system against those of [3]. It can be seen from this figure that the proposed system outperforms Sharma et. al's approach [3] by more than 4% accuracy.

Table 2. Comparing the proposed system against the state-of-the-art system of [3]. Bars number I to V represent, respectively, $(V_{LBP} + T_{LBP})$ with SVM classifier, $(V_{LBP} + T_{LBP})$ with Genetic Algorithm SVM classifier (GASVM), $(V_{LBP} + T_{HDTP})$ with SVM classifier, $(V_{LBP} + T_{HDTP})$ with GASVM classifier, and the proposed system.

methods	I	II	III	IV	V
Accuracy	61%	79%	76%	85%	89%

IV. Conclusion

Stress recognition using computer vision techniques is of great importance as it does not need any contact with users, which is unavoidable in traditional methods. To this end, in this paper we proposed a system that uses facial images of different modalities, including RGB and thermal to make a decision if a user in a current frame is in a stressful situation or not. From RGB and thermal modalities, LBP and temperature of super-pixels have been used as features that are fed to a SVM classifier. The SVM results of the two different modalities are then combined using a score level fusion. The experimental results, showed that the purposed fusion results in a system that outperform the state-of-the-art stress recognition system.

References

[1] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A real-time human stress monitoring system using dynamic bayesian network," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 70–70.

[2] F. Bousefsaf, C. Maaoui, and A. Pruski, "Remote assessment of the heart rate variability to detect mental stress," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*. IEEE, 2013, pp. 348–351.

[3] N. Sharma, A. Dhall, T. Gedeon, and R. Goecke, "Thermal spatio-temporal data for stress recognition," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, 2014. [Online]. Available: http://dx.doi.org/10.1186/1687-5281-2014-28

[4] S. Lupien, F. Maheu, M. Tu, A. Fiocco, and T. Schramek, "The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition," *Brain and Cognition*, vol. 65, no. 3, pp. 209 – 237, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0278262607000322

[5] T. Chen, P. Yuen, M. Richardson, G. Liu, and Z. She, "Detection of psychological stress using a hyperspectral imaging technique," *Affective Computing, IEEE Transactions on*, vol. 5, no. 4, pp. 391–405, 2014.

[6] D. Metaxas, S. Venkataraman, and C. Vogler, "Image-based stress recognition using a model-based dynamic face tracking system," in *Computational Science-ICCS 2004*. Springer, 2004, pp. 813–821.

[7] D. Dinges, E. McGlinchey, S. Venkataraman, and D. Metaxas, "Optical computer recognition of behavioral stress in space flight," *Habitation International Journal for Human Support Research*, vol. 10, no. 3/4, p. 233, 2006.

[8] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation, space, and environmental medicine*, vol. 76, no. Supplement 1, pp. B172–B182, 2005.

[9] H. Gao, A. Yuce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 5961–5965.

[10] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.

[11] H. K. Ekenel and R. Stiefelhagen, "Local appearance based face recognition using discrete cosine transform," in *13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey*, 2005.

[12] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3515–3522.

[13] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3430–3437.

[14] R. Irani, K. Nasrollahi, and T. B. Moeslund, "Improved pulse detection from head motions using dct," in *International Conference on Computer Vision Theory and Applications*, 2014.

[15] J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 4, pp. 988–998, 2010.

[16] J. Fei, Z. Zhu, and I. Pavlidis, "Imaging breathing rate in the co 2 absorption band," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2005, pp. 700–705.

[17] R. Irani, K. Nasrollahi, and T. B. Moeslund, "Contactless measurement of muscles fatigue by tracking facial feature points in a video," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4181–4185.

[18] I. Pavlidis and J. Levine, "Thermal image analysis for polygraph testing," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 21, no. 6, pp. 56–64, 2002.

[19] I. Pavlidis, N. L. Eberhardt, and J. A. Levine, "Human behaviour: Seeing through the face of deception," *Nature*, vol. 415, no. 6867, pp. 35–35, 2002.

[20] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal imaging of physiological stress and its affective potential," *Affective Computing, IEEE Transactions on*, vol. 3, no. 3, pp. 366–378, 2012.

[21] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[22] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep., 2007.

[23] P. Neubert and P. Protzel, "Superpixel benchmark and comparison," in *Proc. Forum Bildverarbeitung*, 2012, pp. 1–12.

[24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2274–2282, 2012.

[25] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 12, pp. 2290–2297, 2009.

[26] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2097–2104.

[27] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," *Trans. on PAMI*, vol. 31, no. 12, pp. 2209–2297, 2009.

[28] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, "Local binary patterns for still images," in *Computer Vision Using Local Binary Patterns*. Springer, 2011, pp. 13–47.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.

[30] Template matching using correlation coefficients. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/28590-template-matching-using-correlation-coefficients

[31] A. Hadid, M. Pietikäinen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–797.

[32] X. Feng, A. Hadid, and M. Pietikäinen, "A coarse-to-fine classification scheme for facial expression recognition," in *Image Analysis and Recognition*. Springer, 2004, pp. 668–675.

[33] M. Taini, G. Zhao, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared video sequences," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[34] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," in *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 2003, pp. 258–264.