ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.3

Facial Expression Recognition for Stress Management to Annotate Music

S.Madhan[1], A.Ramya[2], N.Barani[3], S.Sowndharya[4]
AssistantProfessor[1], Student[2, 3, 4]
Department of Computer Science and Engineering
University College of Engineering, Thirukkuvalai, India

Abstract:
Facial expression is an operative way for humans to interconnect since it contains perilous and necessary information concerning human affective states. It is a critical part of affective computing systems that aim to identify and therefore better rejoin to human emotions. Automatic recognition of facial expressions can be an important component in human - machine interfaces, human emotion analysis, and decision making. However, the task of automatically recognizing various facial expressions challenging. As a result, facial expression recognition has become a prominent research topic in human-computer interaction, as well as the fields of image processing pattern recognition, machine learning, and human recognition. In this project, we will implement the techniques to automatically detect facial parts using HAAR CASCADES algorithm and classify the emotions using Support Vector Machine algorithm. And present playlist of songs which is suitable for his current mood using K-Nearest Neighbor classification algorithm. In testing side, would supply a test image whose expression it desires to recognize. This test image would be matched with facial databases to play music based on recognized emotions. Finally provide emotion based music player with improved recognition rate

Index Terms: Facial image acquit ion, pre-processing, facial feature extraction

1. INTRODUCTION

The topic of stress currently attracts significant attention, not only in research but on social life in general. The public is aware of these phenomena and of its consequences at many levels (e.g. psychological, physical, social, well-being). On the other hand, researchers in many different fields work to find new ways to assess, monitor and reduce stress that can not only answer the interest of the public but also allow a better understanding of the phenomenon. Of all the important perspectives on stress, a particularly interesting one concerns occupational stress. While occupational stress affects individuals at a personal level, there is a special interest in the effects at the organizational level, mainly its economic impact. There is a broad consensus that job stress has a significant economic impact, amounting to billions of dollars each year in the United States alone .These losses are due to the increased cost of medical insurance, excess of pressure on medical facilities and pro-fissional, lower productivity, human error, absenteeism, and so forth. This calls for the development and implementation of initiatives for stress management that can not only reduce these costs but, at the same time, improve well-being, work-place quality, among other indicators. Specifically, we focus on methods that can be used continuously throughout the day in milieus such as the workplace. We seek ways to measure stress over long periods of time that do not influence the workers' routines. Although we address methods that can be deemed as more traditional (e.g. physiological sensors, questionnaires) in the sense that they have been in use for decades, we focus especially on novel methods that, for their characteristics, raise significant interest. Moreover, given their novelty, these methods pose new challenges at several levels (e.g. techno-logical, ethical) and caution is advised when using them. Specifically, we focus on methods that can be used in line with Ambient Intelligence systems, allowing a continuous monitoring of the users while they perform their daily activities, without interference. In this sense, it is important to start by clarifying two concepts that are often found in research in this field that, although different, are frequently used interchangeably: invasive and intrusive. In a physiological sense, an action is called invasive if it infiltrates, cuts or destroys healthy tissue, namely the skin. An intrusive action, on the other hand, is one that intrudes or interferes in one's space, resulting in (often unwanted) changes in routines. Consequently, a non-invasive approach is one in which there is no invasion of the user's body. This includes most of the sensors currently used for stress monitoring (e.g. skin temperature, heart rate). A non-intrusive approach, on the other hand, must meet more strict criteria. Specifically, it cannot change, in any way, the routine of the user. This means that users must be able to carry out their daily activities as if they were not being monitored. This includes, by definition, approaches based on computer vision or speech analysis, for example. Nonetheless, other types of intrusion may be present, as will be addressed later (e.g. the use of a video camera may be seen as a privacy intrusion). These approaches will thus be compared, namely in terms of their degree of intrusion, so that researchers or practitioners can decide on the best to use for each domain of application. The paper is organized as follows. Section 2 addresses the fields of Ambient Intelligence and Ambient Assisted Living, describing their main characteristics and aims. Sec-ton 3 describes stress, its origin, its effects at several levels and its importance, especially in the workplace. Section discourses outdated slants for stress duty, namely those based on physiological sensors and questioners'. Section 5 contains the core of the paper, detailing several new methods for stress assessment that present significant advantages when compared to more traditional ones, making them more suitable to be used in the work-place. These methods are critically analyzed and compared in Section 6, allowingpractitioners to decide on the best method to use in each specific domain. Finally, Section 7 presents a discussion of the main conclusions of this work and points out future research trends and directions.

2. FACIAL IMAGE ACQUISITION

In this module, we capture the face image or upload the datasets. The uploaded datasets contains 2D face images. In face registration we can identify the faces which are captured by web camera. Then web camera images known as 2D images. Mobile phones, tablets, PCs, smart watches which can track several health parameters, smart toys with ingenious independent behavior, houses which can control comfort and safety aspects of our life, smart classrooms, smart offices, automated farming, autonomous cars, autonomous airplanes, and a myriad of other advances have brought computing very close to our daily lives in a way most did not anticipated just a couple of decades ago and it was the topic of science fiction films. This transformation was not simultaneous: the most influential work in this direction is acknowledged to have started within the areas of "Ubiquitous Computing" and "Pervasive computing" [6]. It then progressed into other concepts with "Ambient Intelligence" [7] and "Intelligent Environments" [8].All these can be described as attempts to create "...digital environments that proactively, but sensibly, support people in their daily lives." [9]. Some of these areas put more emphases in different aspects of the system as developers were gaining experience and understanding of the most challenging aspects of these multidisciplinary systems which brought together sensing, networking, human computer in-traction, artificial intelligence, and software engineering, to mention some of the most relevant disciplines. Key to the success of all these systems is that to please the intended users, the system has to have an understanding of the context where services have to be delivered. The subtler this understanding is, the more informed the system to satisfy a given user. This includes understanding personal things, like the preferences or the emotional state of a user. Say I like to be greeted with some ambient music when I arrive home. The system needs to know that I prefer music from J.S. Bach to music from Iron Maiden, but not any music from J.S. Bach will do every day, so if some days I am in need of more cheerful music then perhaps a cantata may be a good choice and if one day I am in need of more relaxing music I may prefer some pieces from the Well-Tempered Clavier. Some days I may not want any music at all. How is the system going to know that? See section on "Mindreading" in [8]. The example above may not seem too important as its related to leisure. However, one of the most important possible applications and one of the most widely researched and tried benefit expected from this area is what is often referred to as Ambient Assisted Living (AAL). "AAL refers to intelligent systems of assistance for a better, healthier and safer life in the preferred living environment and covers concepts, products and services that interlink and improve new technologies and the social environment." [10]. There are several definitions of AAL. However, most of them put emphasis on the safety, health, and well-being of individ-uals. Although these type of benefits are usually placed in the home environment, AAL system do not have to be restricted to houses and can actually redelivered in other places such as the work place, where many people spend a considerable part of their lives. AAL services are also most often associated with older people and in particular with senior citizens experiencing some category of dementia. Although it is true those are the type of applications which have most funding so far, hence more interest, it is clear AAL can help citizens with other conditions, Parkinson's disease, Down's syndrome, autism, etc. From this we can also state that AAL benefits are not only for senior citizens but it is a type of service with the capacity to improve the quality of life of all citizens. Having introduced AAL as a kind of specific branch of Ambient Intelligence with specific interest in the welfare of citizens, it is clear that a system given such a responsibility has to have substantial capabilities to understand what a person is going through at a given time as well as powerful decision-making. For a system to be capable of looking after the welfare of an individual, it has to understand that individual deeply. It is not only a matter of knowing about the preferences of that individual and how those preferences are linked to different situations but it also implies being capable to understand how a user feels 'now'. We can revisit the ambient music scenario, but now imagine the person in question is depressed. If the ambient music is the wrong one for the mood of the user, it may have a detrimental effect, increasing the levels of anxiety, depression or stress of the individual in question. Stress may lead to wrong decisions, which in turn can have undesired consequences resulting in more stress [11].If understanding a specific mental state of an individual like feeling stressed is so important for the success of AAL, how can we do it? There are different approaches, some of them more behavioral and others more biological. By these we mean that some answers to the challenge try to under-stand how the individual is behaving, e.g., body language, whilst the latter approach relies more on measuring specific personal body parameters which can provide an indicator, e.g. high blood pressure as a potential indicator of stress. The next sections of this paper provide a more specific account of these approaches and highlight the challenges behind each of these options

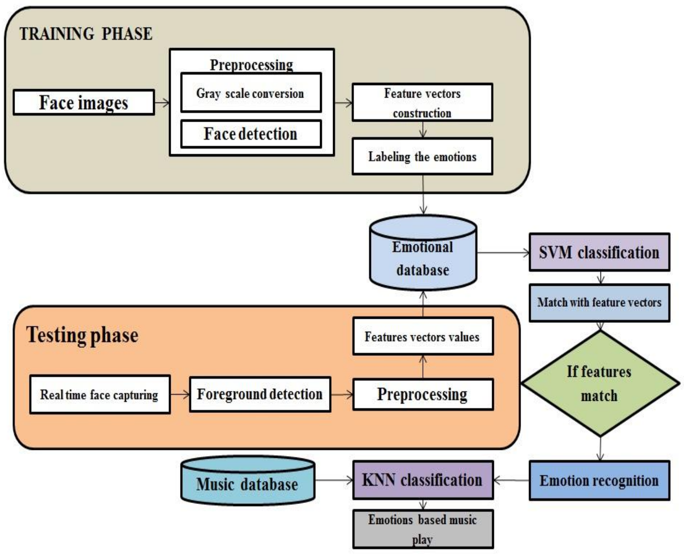

Figure.1.Architecture diagram.

3. PREPROCESSING

Perform the preprocessing steps such as gray scale conversion, invert, and border analysis, detect edges and region identification. Gray scale images are also called monoc hromatic, denoting the presence of only one (mono) color (chrome).And also remove the noises from face images.

4. FUNDAMENTAL CONCEPTS

In modern science, stress started to be studied at physiological level, in the decade of1950.This resulted in a set of reliable physiological indicators for the study of stress that supported the development of the bio-feedback units available nowadays. In the 70's researchers started studying the somatic disorders resulting from these biologic aspects [14]. At the same time, Hans Sale provided an accurate and simultaneously accessible definition of stress [15], putting forward the notion of stressor and addressing the hormonal changes caused by stress .Although such views have changed throughout history, there is anagreement that responses to stress are cord- noted by a so-called stress system, whose composition is nowadays well studied and known to include as main components the corticotrophin-releasing hormone and locus cerulean-norepinephrine/autonomic systems and their peripheral effectors . Moreover, the things of stress at changed levels (e.g. behavioral, peripheral, physiological, and cognitive) are nowadays becoming known. As a collusion, an up-to-date view of stress looks at it as a physic-physiologic arousal response occurring in the body as result of stimuli. A single-modality approach for measuring the effects of stress would thus not be suited, as some experimental re-salts demonstrate [17].In fact, for a sufficiently precise and accurate measurement of stress, a multi-modal approach must be considered. The diagram depicted in Figure 1 represents a simplified multi-modal view on stress as considered in this paper. This diagram is composed of two main parts: the upper part concerns the predictive aspects of stress while the lower part concerns the diagnostic aspects. The Predictive part of the model considers the following aspects: Context, Profile, Goal and Trait. Context includes meaningful information to describe the different dimensions of the individual, including the historical, economic, social or geographical contexts. Numerous studies exist that map such information to a base level of stress: the effect of socioeconomic status [18], social or geographical context [19], [20], [21] or individual economic situation [22], just to name a few.

The Profile of the individual includes personal information and characteristics that have an ongoing influence on the level of stress. These include age, gender, marital status, number of dependents [23], type (or lack) of employment [24], job category, among others. The Goal of the individual at a given moment in time or, likewise, the objectives, aspirations or ambitions also has a significant influence on the level of stress. Namely, individuals with higher ambitions are generally known to be under increased stress, resulting from the continued effort of trying to achieve above average standards [25].Finally, Trait is related to the personality of each individual, i.e., habitual patterns of behavior, thought or emotion. Some traits are more generally associated with stress than others [26]. As an example, an impulsive individual is generally a more stressed one, with stress driving his hasty decisions. In the diagnostic part of the model, a larger number of components could be included. Namely those oriented towards psychological or psychosomatic diagnostics, i.e., subjective self report mechanisms such as surveys or questionnaires .We however focus on objective measures rather than the subjective ones, especially those that can be used to provide real-time feedback. Thus, the Diagnostic compo-nets of the model include Physical, Physiological, Behavioral and Performance aspects. Physical aspects include, in a general way, bodymove-ments or postures that may have some particular meaning in terms of stress assessment. Especially interesting are aspects such as eyelid movement, facial expressions, body movements (e.g. specific gestures, head movements, repetitive movement patterns) or pupil movement and dilatation Physiological diagnosis aspects are those that provide the most reliable diagnose of stress. In fact, many approaches exist nowadays that can evaluate the level of stress of an individual from physiological indicators with signify-cant precision, as will be addressed in detail in Section 4.On the other hand, the behavior of an individual can be seen as the visible end of his inner self. In that sense, aside from other aspects, behaviors (and especially changes in behaviors) may also be a good indicator of stress effects. Given the scope of this paper, particular attention will be dedicated to behaviors when interacting with technological devices or behaviors that can be acquired within technology ice environments, non-intrusively. Finally, the Performance of an individual is significantly affected by stress. The optimum level of stress will maximize performance. A higher level of stress may increase performance temporarily but will soon wear the the individual. A lower level of stress will decrease productivity and lead to increasing lethargy. Thus, tests that evaluate performance in given tasks, for which standard performance measurements are known, can be a good indicator of the effects of stress on the individual. From a high level point of view, two different types of stress can also be identified: acute and chronic stress. Acute stress comes from recently acknowledged demands and pressures and from anticipated demands in the near future. On the other hand, chronic stress is long-term, due to social or health conditions, dysfunctional families, among many other issues. This type of stress will have nefarious effects on the body and mind of the individual, slowly wearing him away day after day. Acute stress, because it is short-term, won't do the extensive damage associated with chronic stress, although overtime frequent acute stress may contribute to the development of chronic stress. Never the-less, it will instantaneously influence the performance of the actions being carried out. Given the broadness of the field, in this paper we clearly focus on acute cognitive stress. Indeed, most if not all non-intrusive and non-invasive current methods for stress assessment are based on the observation of changes on the individual (as detailed in Section 5). When considering acute stress, these changes are easily observed as they constitute significant deviations from an otherwise regular behavior or state [27]. Chronic stress, on the other hand, is more difficult to detect using these means as the individual is constantly experiencing the effects of stress, thus no abrupt changes are observed [28]. We do not mean to imply that it would be impossible to accomplish. However, given the character is-tics of chronic stress, its detection using the means explored in this paper would require more extensive data collection about each individual, spanning longer time-frames. Once the face location has determined and the face extracted, the system moves onto extracting regions of interest, in particular, the eyes, eyebrows and mouth.

This stage is followed by a feature extraction stage, where various algorithms are used to implement multiple techniques. Their results are evaluated and presented in the later Testing and Evaluations. This extracted data is fed to a classifier which determines the emotion of the user. The system uses two particular algorithms for face detection, hear features using the Viola-Jones Technique and Local Binary Patterns. The report does not delve too deeply into these, as they are not central to the project. Thus the denotes emotion to either the stress expression in face.

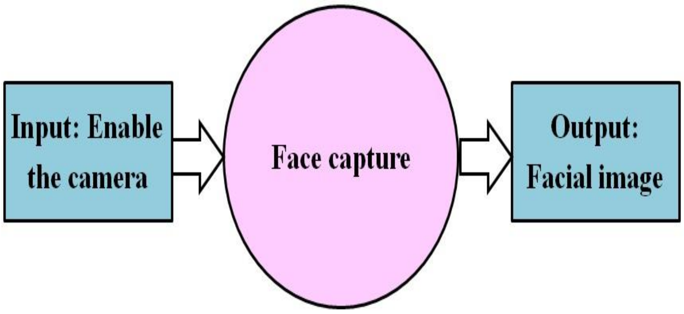

Figure.2. facial image acquisition

In this module, we capture the face image or upload the datasets. The uploaded datasets contains 2D face images. In face registration we can identify the faces which are captured by web camera. Then web camera images known as 2D images. In this module, perform the preprocessing steps such as gray scale conversion, invert, and border analysis, detect edges and region identification. The Gray scale images are also called unicolor, denoting the occurrence of only one color. The edge detection is used to analyze the connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Numerous tactics have been proposed to classify human moving states. The features used are typically based on displacements of specific points or spatial locations of particular points; this technique is known as Facial Action Coding System (FACS). In an approach taken by Liu et al in [3], he presents an algorithm for classification of brain electrical signals in human emotions. This algorithm was based on the model of fractal dimension. He proposed a bi dimensional Valence - Arousal approach, where by the six emotions are divided into different categories. Black et al. explored the use of local parameterized replicas of image motion for improving and identifying the non-rigid and spoken motion of human faces. They used these parametric models to extract the shape of the mouth, eyes and eyebrows. They realized a high success amount of 95% to order Happy, 90% to classify Anger and a 100% success rate to classify the sad emotions. On the other hand the approach used by Jacob and Davis [2] in which facial expressions are recognized in image sequences using statistical properties of the optical flow with only very weak models of facial shape. In this project several approaches are considered, including a Principal Component Analysis (PCA) approach, using multiple Facial Action Units and different feature extractors with clustering approaches. Each of these approaches is used with different classifiers to determine the emotion of the user with the accuracy tested on a data set of 40 subjects each. The system consists of two major sections: Image analysis; and music file analysis. The program heavily focuses of image analysis using various training samples to train and predict data. It consists of multiple classifiers; various feature extractors and an evaluation function. Past works and research heavily influenced the approach considered towards development for this project. The key features include; a live learning algorithm for associating each music genre with an emotion based on the preference of the user, recording pictures of the user based on a predefined time scale to capture changes in emotion and organizing music playlists in a lexicographic order in the end to output a music playlist.

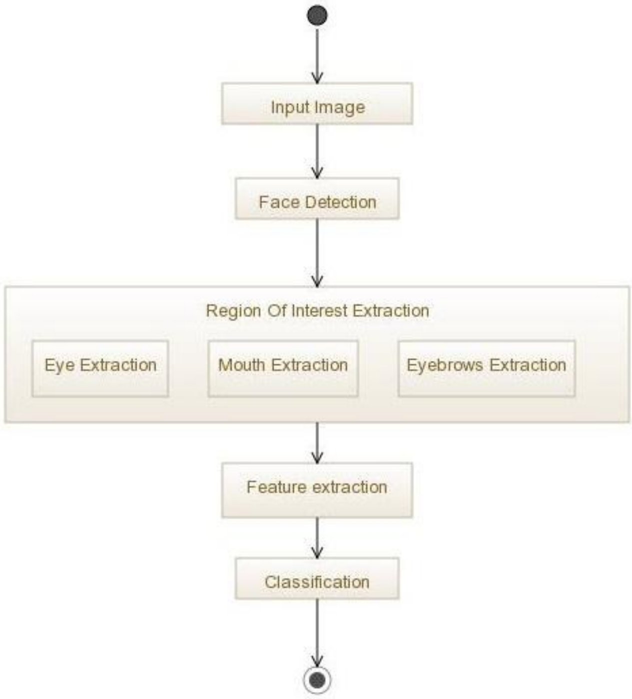

Figure.3. overview of the emotion recognition system

This algorithm has its origins in the 2D touch analysis of images. The basic idea is to abridge the local structure of an image by equating each pixel with its national. The proposed idea is to take a pixel in the center and threshold it against its neighbors. If the value is less than its neighbor, denote it with a 0 or else denote it with a 1. Hence you end with a binary value for each pixel, known as the Local Binary Patterns (LBP). A useful extension to the original operator is the so-called uniform patterns [8]. These patterns improved the efficiency of the system by reducing the length of the feature vector and hence implementing a simple rotation invariant descriptor. The system makes use of these uniform patterns to determine face locations from a camera feed.

- Eye Width = (Right Eye Width + Left Eye Width) / 2 = ((d1 - d2) + (c2 - c1)) /2
- Eye Height = (Right Eye Height + Left Eye Height) / 2 = ((d4 - d3) + (c4-c3))/ 2
- Mouth width = f2 - f1
- Mouth Height = f4 - f3
- Distance between eyes = d2 - c2

The calculation steps as follows

Feature = $w_1 * \mathrm{RecSum}(r_1) + w_2 * \mathrm{RecSum}(r_2)$

Weights can be positive or negative
Weights are directly proportional to the area
Calculated at every point and scale

It includes weak classifier such as

A weak classifier ($h(x,f,p,\vartheta)$)
Consists of
-feature (f)-threshold (ϑ)
-polarity(p), such that

$$h(x,f,p,\vartheta) = \begin{cases} 1 & \text{if} pf(x) < p\vartheta \\ 0 & \text{otherwise} \end{cases}$$

Requirement

Should perform better than the random chance

Cascade creation algorithm as follows:

$f_0 = 1$
i=0
while $f_i > t_{arget}$ and i<n Stages
i=i=1
Train classifier for stage i
Initialize the weights
Normalize the weights
Pick the (next) best weak classifier
Update weights
Evaluate f_i
if $f_i > f$
go back from normalize the weights
Combine weak classifiers to form the strong stage classifier
Evaluate F_i

5. CONCLUSION

Real time facial datasets are used implemented the system to recognize the emotions classify the music database with improved accuracy rate. In summary it is clear to see that the project set out and achieved its basic functionality. The feature extraction techniques could evidently be improved and much more advanced techniques exist to improve the accuracy of an emotion recognition system. Within this project there is scope for more work that could allow for the complexity of the modeling to be increased, which could evolve the result of the binding between the two main domains of current research: Music theory and Image analysis. The biggest limitation of the project is in the general approach taken for emotion recognition using facial cues. The camera must take a full frontal image of the user to determine the emotion of the user accurately. Even though rotation invariance is taken into account for two of the feature extractors, accuracy rates of subjects facing the camera sideways was extremely low in comparison to frontal images. Indeed SIFT depicted an average accuracy of 63% whilst SURF fared slightly better with an average accuracy rate of 68 %. Another massive limitation faced during the testing and development of the project was a relatively small sized training dataset. For ethical reasons express permission is required to use a database of faces depicting various emotions. Despite constant efforts only the Yale Face Database granted permission. Hence forth a small training data set was used which resulted in a relatively low accuracy rate for the system. To develop a music information retrieval system, initially the library *Maryssa* (Music Analysis, Retrieval and Synthesis for Audio Signals) was determined to be appropriate. The library turned out to be unstable and within a few weeks of development, it was unfeasible to use this with windows and Opens. Hence the project shifted from a machine learning approach to a Meta data extraction for music classification. This was another hindrance and limitation faced during the project development. This chapter discusses the conclusions and limitations of the project. The project set out to investigate the impact of various techniques in classifying human emotion. It also set out to use the classified emotion to sort out music playlists for a person. This chapter also performs an evaluation on the approach taken. The lessons learnt during the course of this project are also discussed and finally the chapter concludes with the possible future directions of the project. During the development of the project, existing techniques for feature extraction and emotion recognition were thoroughly researched, highlighting the benefits and problems with each associated. After thorough research 3 feature extraction techniques were decided upon and further techniques to improve classification accuracy were chosen. Further research dictated the use of the Support Vector Machines as the classifier to be used. The evaluation of the project indicates that the preliminary objectives of the project have been met.

The key objectives were:

1. Implement basic feature extraction algorithms (SIFT and SURF) to determine their impacts on an emotion recognition system.
2. Extract information from music file to sort and classify a music playlist based on human emotion

Added functionality implemented for the project included using various techniques to improve the accuracy of the classifier. The two proposed models: PCA and Bag of Words model were implemented and their relative impacts were analyzed. Thorough testing showed that the use of these models vastly improved accuracy, the use of PCA in particular produced the highest accuracy rates in emotion classification. These chapter further details the limitations of the project, the issues faced and the lessons learnt during the project development. Further testing and research suggests that each and every emotion in the extreme could be confused to be the neutral emotion. Hence forth with each emotion and each technique, some of the emotions have always been misclassified as neutral. This occurs when a person has a slight smile and a barely noticeable frown etc. Hence only determining emotions using just facial cues might not be the most viable approach. Other features such as the hand movement or a person's voice need to be taken into account and using all these features, a mind is able to accurately perceive the human emotions. In terms of the"goodness" of each method, it is impossible to point out, in absolute terms, the best method. In a general way, the methods that achieve the highest score are those based on the mouse and the keyboard. These methods also achieve high scores in characteristics such as cost-effectiveness, low intrusiveness or privacy. However, if the focus of the organization is on accuracy, methods such as computer vision or wearable's should be considered instead (or in addition). These methods are, however, also those that represent the higher cost.

6. REFERENCES

[1]. M. Liu, S. Shan, R. Wang, and X. Chen, "Learning expressionless on patio-temporal manifold for dynamic facial expression recognition," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2014.

[2].H. Jung, S. Lee, J. Yam, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in The IEEE International Conference on Computer Vision (ICCV), IEEE, 2015.

[3]. Y. Sun, X. Wang, and X. Tang, "Deep convolutional network waterfall for facial point finding," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013.

[4]. L. Hong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012.

[5]. X. Xiong and F. De la Torre, “Overseen descent method and its applications to face position,” in The IEEE Discussion on Computer Vision and Pattern Gratitude, IEEE, 2013.

[6]. T. A. Beehr and R. S. Bhagat, Human stress and cognition in organizations: An integrated perspective.Wiley Interscience, 1985.

[7]. T. A. Beehr, Psychological Stress in the Workplace (Psychology Revivals). Rutledge, 2014.

[8]. D. Carneiro, P. Novais, R. Costa, P. Gomes, and J. Neves, “Emon: embodied monitorization,” in European Conference on Ambient in Intelligence. Springer, 2009, pp. 133-142.

[9]. S. Jones, S. Hara, and J. C. Augusto, “efriend: an ethical framework for intelligent environments development,” Ethics and InformationTechnology, vol. 17, no. 1, pp. 11-25, 2015.

[10]. D. Carneiro, A. Pimenta, S. Gonçalves, J. Neves, and P. Novais, “Monitoring and improving performance in human-computerinteraction,” Concurrency and Computation: Practice and Experience, 2015.

[11]. M. Weiser, “The computer for the 21st century,” SIGMOBILE Mob. Comput. Commun. Rev., vol. 3, no. 3, pp. 3-11, Jul. 1999. [Online].

[12]. Available: http://doi.acm.org/10.1145/329124.329126 E. Aarts, R. Harwig, and S. M, Ambient Intelligence. McGraw-Hill, Inc., 2001.

[13]. J. C. Augusto, V. Callaghan, D. Cook, A. Kameas, and I. Satoh, “Intelligent environments: a manifesto,” Human-Centric Computing and Information Sciences, vol. 3, no. 1, pp. 1-18, 2013.

[14]. J. C. Augusto, “Ambient intelligence: Basic concepts and applications,” in Software and Data Technologies. Springer, 2006, pp. 16-26.

[15]. G. van den Broek, F. Cavallo, and C. Wehrmann, AALIANCEambient assisted living roadmap. IOS press, 2010, vol. 6.