Towards In Situ Affect Detection in Mobile Devices: A Multimodal Approach

Mohammad Adibuzzaman[1], Niharika Jain[1], Nicholas Steinhafel[1], Munir Haque[2], Ferdaus Ahmed[1], Shiekh Iqbal Ahamed[1], Richard Love[3]
[1]Department of Math, Statistics and Computer Science
Marquette University, Milwaukee, WI, USA
[2]Department of Computer and Information Sciences
University of Alabama at Birmingham, Alabama, USA
[3]International Breast Cancer Research Foundation, Madison, WI, USA
[1]{mohammad.adibuzzaman, niharika.jain, nicholas.steinhafel, ferdaus.kawsar,sheikh.ahamed}@marquette.edu,
[2]mhaque@uab.edu,[3]richard@ibcrf.org

ABSTRACT

Most of the research in multi-modal affect detection has been done in laboratory environment. Little work has been done for in situ affect detection. In this paper, we investigate affect detection in natural environment using sensors available in smart phones. We use facial expression and energy expenditure of a person to classify a person's affective state by continuously capturing fine grained accelerometer data for energy and camera image for facial expression and measure the performance of the system. We have deployed our system in natural environment and have provided special attention on annotation for the training data validating the 'ground truth'. We have found important correlation between facial image and energy which validates Russell's two dimensional theory of emotion using arousal and valence space. In this paper, we have presented initial findings in multi-modal affect detection.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine System; I.4 [**Image Processing and Computer Vision**]: Feature Measurement

General Terms

Performance

Keywords

Algorithms, Experimentation, Human Factors

1. INTRODUCTION

Affect sensitive applications are being developed in a number of domains which include learning technologies, autism spectrum, gaming, robotics and Human Computer Interaction [6]. Though there has been significant research in the field of affective computing, most of the research is done for unimodal affect detection. Interestingly, human affective state is never a unimodal expression. Any affective state such as happiness, sorrow, anger etc. almost always involves two or more of the modalities such as facial expression, body movement, speech data and other emotional cues. Thus a multimodal affect detection system may have better accuracy and reliability that has largely been ignored in the literature.

In the field of affective computing, multimodal real time implementation is widely advocated but rarely implemented [6]. Research has been done for affect detection from facial expression, speech data, body gesture, heart rate, skin conductance, pressure sensor and other inputs. Research has been conducted for affect detection using all of these communication channels in a laboratory environment. In addition to facial expression research work inspired by Ekman [19], there are other approaches such as region based or holistic approach. Instead of different Action Units (AU) proposed by Ekman, region based approach uses certain regions of the face such as eye, eyebrow and mouth. Some of the expression recognition system uses various pattern recognition techniques [12] and geometric and appearance feature-based methods [22].

Beyond facial expression, several affective computing applications focus on detecting affect by using machine learning techniques to identify patterns in physiological activity [23]. These patterns in physiological activity correspond to different emotions. Previous research was done using facial expression and speech data fusion. In the literature we have found only one work with the fusion of multiple modalities which includes physiological activity [7]. A correlation between automated assessment of mental (or physical health) and the result of the gold-standard surveys with sensor based measurements were found in [15].

There has been much research for multi-modal affect detection in the past decade [23]; but few research studies consider hand-held devices which are equipped with sensors that can be used for affect detection. The affect detection technique for hand-held devices will not require a multitude of the complicated sensors the user has to wear in laboratory environments. Rather, we need a solution that uses the existing sensors of the hand-held devices.

Fortunately, smart phones are equipped with many sensors such as camera, microphone, accelerometer and GPS. Each of these sensors can be used for capturing affective state in different channels like facial expression, speech, activity and context. Further-

RACS'13 October 1-4, 2013, Montreal, QC, Canada.

more, the power of hand-held and mobile devices is increasing at a tremendous speed. With the advent of mobile technology, many smart phones now have an accelerometer along with a built-in camera. These two sensor devices are critical for automatic affect detection. Accelerometer data could be used for estimation of how much energy a person has exerted over a period of time. Physiopsychology research shows that there is a considerable correlation between energy expenditure and affective states.

Therefore, we choose to use facial expression and physiological activity for multimodal affect detection in natural environment using camera and accelerometer of smart phones using Naïve Bayes fusion. We evaluated the system performance and found significant improvement of our system, which also include energy expenditure, over unimodal system which uses only facial expression.

1.1 Contributions

The ability to detect and understand human affective state is at the core of human intelligence which is indispensable from human behaviour. This is also true to understand human decision making process as well as consumer behaviour [3]. In this paper, we present an automated system that can detect affective state in natural everyday setting using sensors available in smart phones. The contribution of this work includes:

- An automated system that can be used in day to day life without any interference to the user.
- This work shows that the system performs better when we do the fusion of facial expression and energy than only from facial expression.

2. STATE OF THE ART

One of the major problems experienced while doing study on affective computing was to correctly identify the human emotions in a laboratory setting [14]. This is because, in contrary to the natural world scenarios, the controlled environment does not offer the natural occurrence of actions and corresponding reactions. Moreover, the laboratory proceedings do not take into account the results of other concurrent activities happening in day-to-day life. In addition to this, the identification of emotions is sometimes based on the reports created by self-analysis and hence suffers from its own set of limitations.

To overcome these factors, a 7-day study was conducted in [4], where data related to the affective states were collected through several means. First source being the 'in situ' analysis taking into affect the elimination of bias and independence of axes. Second source of data was the overall 'end-of-day' rating and third source accounted the scores given by the third-party raters. The data collected from these different sources was then triangulated so as to achieve a data set containing mutually agreeing information. This information when fed to a J48 decision tree returned highly accurate results (100% accuracy) in terms of high or low activation states.

For modeling of affective data in natural environment, a study was conducted by [5] so as to collect the sensor data corresponding to physical activity, heart rate and galvanic skin response. The data so obtained was then aggregated and stored in a mobile device which was also used as a user interface. This mobile device was also used to capture audio data related to the subjects involved. The study concluded that the selection of correct time window and having a customized window around annotated events played key roles in obtaining the correct emotion analysis.

The power of mobile devices has been further explored for affect annotation by incorporating the multi-modal technique for assessing physical as well as mental well-being [15]. A study was conducted wherein the subjects were given a device consisting of various sensors which could help collecting data required for the above mentioned assessment. The classification of collected data as speech and activity was done using two-state hidden Markov Models (HMM) and decision-stump classifiers respectively. A correlation between automated assessment of mental, and/or physical health and the result of gold-standard surveys was found so as to stress upon the accuracy of sensor-based measurements.

Healey et al. tries to standardize the affective data annotation in [4] and [5]. But their approach used sensors not only that are available in mobile devices, but also external sensors. Again, a comprehensive study about the classification algorithm was not present. [15] shows the correlation of gold standard surveys with sensor data capture, but that does not provide a study only for mobile device. The participants had to wear other sensor devices for affective data annotation. In our study we overcome both of these problem by using only one smart phone for the user as well as a comparative study of the result about multimodal versus unimodal system. All of these works provide the techniques and results of using different modalities. But none of the research study uses only smart phones for collecting data. Also, a comparative study between unimodal system versus multimodal system was not present.

3. OUR APPROACH

To capture the arousal and valence space, we used facial expression and energy exertion of a person. From the field of psychology, arousal space can be captured by heart rate, pupil size or energy expenditure; all of them can be captured from the sensors available in smart phones. In this paper, we used facial expression for valence and energy expenditure for the arousal space. We used eigenface algorithm for detecting facial expression and later used mean of different affective states as the second feature for multimodal affect detection using Naïve Bayes fusion. In this section we describe the methods we used for the annotation of in situ affective data and the reason for using facial expression and energy expenditure. In the next section, we describe the details of our classifier.

3.1 Selecting Modalities

Emotion labeling is moderately less work in laboratory environment where the researcher can control the environment, recreate the situation, recording can be done accurately and the person can be interviewed for his/her annotations. However, the emotion labeling can still be flawed since in controlled environment people might act differently, both physiologically and cognitively. In situ capturing of affective state captures the natural data, but it needs more methodical approach for emotion journaling. The participants need to be trained well and the data labeled needs to be verified later. Our goal is to minimize the error for establishing the 'ground truth', which defines true affective state for the given data in machine learning algorithm for classification.

According to the Russell's circumplex model of emotion, each affective state can be represented in 2D space [18]. The horizontal axis represents the valence and the vertical axis represents the arousal space. Valence represents how good or bad a person is feeling, and arousal represents how much a person is aroused. Therefore, we hypothesize that if we could capture the arousal space data from accelerometer, we could better classify the affective states. For example, for the happy state, this is a positive feeling and a person might have some kind of excitement. On the other hand, for sad feeling, it is a negative feeling and the person may have less movement, which corresponds to less energy expenditure.

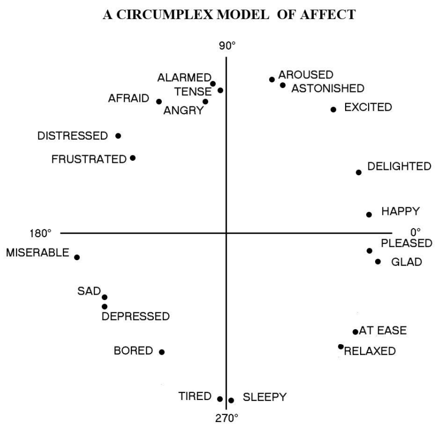
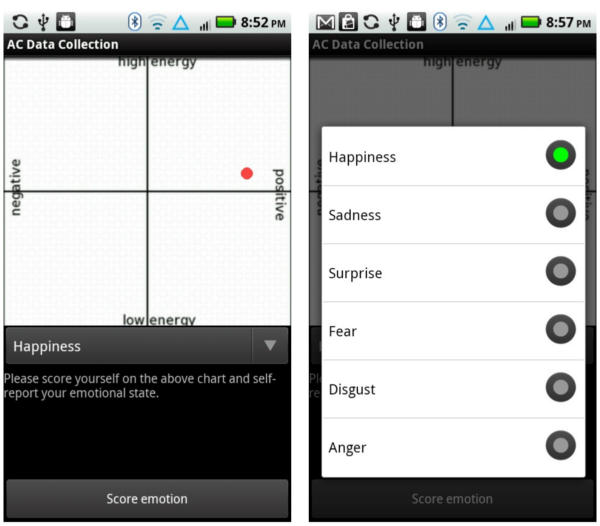


Figure 1: Russell's circumplex model of emotion [18].

3.2 Emotion Journaling

We used smart phones for emotion journaling. Smart phones give us the opportunity for labeling emotion as soon as it occurs with real time sending and storage capability. Eight participants were recruited for the study, aged between 21-33 all of whom are students. Participants were asked to carry the smart phones and annotate the data for at least 5 times a day for a seven day period. The participants will be referred as PA in this work.



(a) (b)

Figure 2: Annotation of emotion data: (a) Annotation of affective state using Russell's 2D emotional space. (b) Annotation of affective state using radio button.

For the journaling of emotional data, we used camera and accelerometer data of smart phones for facial expression and activity. Also location data was stored using GPS of the smart phones that might give us the context information. Three android phones were used, two Droid X with android operating system of 2.2 and one Samsung Galaxy Nexus with android operating system of 3. PAs were asked to take a facial picture with the smart phones and then label the data. The labeling was done using two sources for capturing the natural feeling. One is using the *Mood-map* which corresponds to Russell's circumplex model and the other is radio button from which the user can pick one from the six basic emotions. Figure 2 shows the interface for mood-map as well as the radio buttons. Continuous and fine grained accelerometer data for fifteen minutes before taking the picture and location data were also recorded and then sent to the server using the phones' internet.

3.3 Journaling Training

Each participant was asked to keep the smart phone for one week. Before handing over the smart phone, they were trained on how to use the application for emotional data annotation. During that period, PAs were asked to carry the smart phones six to eight hours a day and label the data whenever any emotional event occurred. They were trained to use the touch based application as well as how to take the picture, use the mood-map and upload the data. Furthermore, there was constant communication between the PAs and the researcher for any question from the participant.

3.4 Algorithm Design

The algorithm for affect detection from facial expression and accelerometer data can be discussed in different components;face detection, affective state from facial image, energy expenditure from body movement and fusion using Naïve Bayes. Each of the components are discussed here.

3.4.1 Face Detection

Pixels corresponding to skin are different from other pixels in an image. [12] has shown the clustering of skin pixels in a specific region for Skin color modelling in chromatic color space. Though the skin color of persons vary widely based on different ethnicity, research [19] shows that the still form a cluster in the chromatic color space. After taking the image of the subject we first crop the image and take only the head portion of the image. Then we use skin color modeling for extracting the required facial portion from the head image.

3.4.2 Affective State from Facial Image

For this part we use a combination of Eigenfaces, Eigeneyes, and Eigenlips methods based on Principal Component Analysis (PCA) [22][23]. This analysis method includes only the characteristic features of the face corresponding to a specific facial expression and leaves other features. This strategy reduces the amount of training sample and helps us make our system computationally inexpensive which is one of our prime goals. These resultant images are used as samples for training Eigenfaces method and M Eigenfaces with highest Eigenvalues. We generate the Eigenspace as follows:

- The first step is to obtain a set S with M face images. Each image is transformed into a vector of size N^2 and placed into the set, $S = \gamma_1, \gamma_2, \gamma_3, ..., \gamma_M$
- Second step is to obtain the mean image ψ

$$\psi = \frac{1}{M}\sum_{n=1}^{M}\gamma_n$$

- We find the difference ψ between the input image ϕ and the mean image,$\phi_i = \gamma_i - \psi$

- Next we seek a set of M orthonormal vectors, μ_M, which best describes the distribution of the data. The k^{th} vector, μ_k, is chosen such that

$$\psi = \frac{1}{M}\sum_{n=1}^{M}(\mu_k^T\phi_n)^2$$

- λ_k is a maximum, subject to

$$\mu_l^T\mu_k = \begin{cases} 1, & \text{if } l == k. \\ 0, & \text{otherwise.} \end{cases}$$

  where μ_k and λ_k are the eigenvalues and eigenvectors of the covariance matrix C.

- The covariance matrix C has been obtained in the following manner

$$\psi = \frac{1}{M}\sum_{n=1}^{M}(\phi_n\phi_n^T)^2 = AA^T$$

  where $A = [\phi_1, \phi_2, \phi_3, ..., \phi_m]$.

- To find eigenvectors from the covariance matrix is a huge computational task. Since $M$is far less than N^2 by N^2, we can construct the M by M matrix,

$$L = A^TA$$

  where $L_{mn} = \phi_m^T\phi_n$

- We find the M Eigenvectors, v_l of L.These vectors (v_l) determine linear combinations of the M training set face images to form the Eigenfaces u_l.

$$\mu_l = \sum_{k=1}^{M}v_{lk}\phi_k$$

  where $l = 1, 2, 3, ..., M$

- After computing the Eigenvectors and Eigenvalues on the covariance matrix of the training images
  - M eigenvectors are sorted by Eigenvalues
  - Top eigenvectors represent Eigenspace

- Project each of the original images into Eigenspace to find a vector of weights representing the contribution of each Eigenface to the reconstruction of the given image.

When detecting a new face, the facial image is projected in the Eigenspace and the Euclidian distance between the new face and all the faces in the Eigenspace is measured. The face that represents the closest distance will be considered as a match for the new image. Similar process is followed for Eigenlips and Eigeneyes methods. The mathematical steps are as follows:

- Any new image is projected into Eigenspace and we find the face-key by
  $\omega_k = \mu_k^T$ and $\omega^T = [\omega_1, \omega_2, \omega_3, ..., \omega_M]$

  where, u_k is the k^{th} eigenvector and ω_k is the k^{th} weight in the weight vector $\omega^T = [\omega_1, \omega_2, \omega_3, ..., \omega_M]$

- The M weights represent the contribution of each respective Eigenfaces. The vector Ω, is taken as the 'face-key' for a face's image projected into Eigenspace.

- We compare any two 'face-keys' by a simple Euclidean distance measure

$$\epsilon = ||\Omega_a - \Omega_b||^2$$

- An acceptance (the two face images match) or rejection (the two images do not match) is determined by applying a threshold.

3.4.3 Energy Expenditure from Body Movement

There exists a significant correlation between accelerometer data and the work done by a person. It is found that the energy measured by ADInstrument Exercise Phsyiology Kit is highly correlated with accelerometer energy when the phone is positioned at the waist [2].

Droid X uses the STMicroelectronics LIS331DL accelerometer. In our study, 2 Droid X 3G devices running Android OS 2.2 and one Samsung Galaxy Nexus with Android OS 3 were used as acceleration measurement platforms.

Since this is a piezo-resistive accelerometer, low pass filtering is required to acquire the true activity-component. We applied low-pass filtering on the raw accelerometer data, as its output includes a DC gravitational contribution. In the literature, the ideal cut-off frequency or the filter ranges from 0.1 Hz to 0.5 Hz. We used 0.5 Hz filter in Matlab to exclude the gravitational contribution. After testing the varying frequency in this range, we found good result preserving the activity contribution.

To correlate accelerometer data with energy expenditure of a person, the accelerometer's three dimensional vector needs to be summarized as one scalar value that represents physical activity intensity over small time periods [2]. This scalar value is considered accelerometer energy spent by the user. To calculate accelerometer energy, several different methods have been proposed, but the most used one is the summation of time integrals of accelerometer output over the three spatial axes [2]. We adopted this method. The accelerometer energy is calculated according to the following formula:

Accelerometer energy $= \int_{t_0}^{t_0+T} |a_x| + |a_y| + |a_z| dt$

Here a_x,a_y and a_z are low-pass filtered accelerometer data corresponding to the x, y, and z axes. For calculating the values of this equation, we found the accelerometer input data on each of the axes. Then low pass filtering was used on each axis input. Next, we calculated the absolute value of the accelerometer inputs and found the integration during fifteen minutes time before taking the user image.

3.4.4 Fusion Using Naïve Bayes

We found the mean of the energy data for different affective states and those means were used as a separate feature for the fusion. Table 1 summarizes the mean of the energy for different affective states. Those means were used as the additional feature for our fusion.

It is argued that human behaviour is close to that predicted by Bayesian decision theory [8]. Different probabilistic graphical model algorithms are used in the literature like Hidden Markov Model (HMM) and Support Vector Machine(SVM).

In our fusion, we used Bayesian classifier. Since we are working only on two modalities, we argue Naïve Bayes algorithm would be a better fit, which performs better with small number of features and potentially large data for fusion . Fusing the modalities of facial expression and energy data at decision level enables us to gain the knowledge about the relationship between these two modalities for a particular affective state [9].

The Bayesian fusion framework that we apply is proposed in [20]. It uses the conditional error distributions of each classifier

Affective State	Energy(mean)
Anger	8.56E+00
Disgust	2.24E+01
Fear	4.12E+01
Happy	1.51E+01
Sad	4.28E+00
Surprise	3.56E+01

Table 1: Mean of energy for different affective states.

to approximate uncertainty about that classifier's decision. The combined decision is the weighted sum of the individual decisions. Given a problem with K classes and C different classifiers, λ_i, $i = 1, ..., C$ we like to infer the true class label ω, given the observation x. Assuming that for each classifier λ_i we have a predicted class label ω_k, where $k = 1, ..., K$ then the true class label can be derived as follows:

$$P(\omega|x) \approx P(\omega|\omega_k, \lambda_i)P(\omega_k|\lambda_i, x)P(\lambda_i|x)$$

Probabilities $P(\omega|\omega_k, \lambda_i$ and $P(\lambda_i|x)$ are used to weight the combined decision and can be approximated from the confusion matrix of classifier λ_i.

We used the energy expenditure data of the same persons from our facial expression database. When the users simulated affective state, their energy data was also collected for the last 15 minutes before taking the photograph.

4. EVALUATION

We evaluated the system in four different ways. Validating the ground truth, performance of unimodal system with only facial expression,validating energy data, and performance of the multimodal system.

4.1 Validating The Ground truth

After the data collection, each day the participants were interviewed and asked about their labeling. We found that some data were not properly labeled. Due to ambiguity of the context, some data were also discarded. For example, on one occasion PA2 said, *'I was feeling very good with my grade, but did not have much movement since I was sitting on my desk. So I labeled the emotion as positive in valence but negative in arousal and did not know which one to pick from radio button. So I selected sad.'* We only incorporated the data that the researcher and the PA we agreed on to be of any particular affective state.

4.2 Unimodal System With Facial Expression

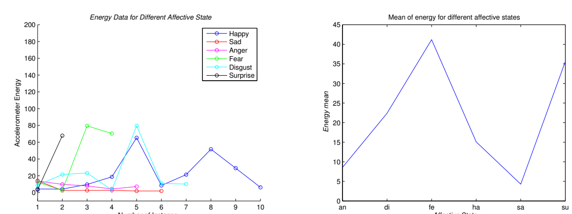
We trained our database with the pictures taken by the camera of the smart phones. Then for each image in the training database, we used our classifier for facial expression and found 89% accuracy. The confusion matrix for facial expression is given in Table 2. We found that the pictures taken by the camera for which the environment was dark, the system gave inaccurate results and the image was not properly classified. We got one inaccurate results for each of the expression anger, sad, disgust, fear and surprise. From the result, we conclude that the classifier works well with the training database as long as the image is taken properly with proper lighting. It does not depend on any particular expression.

a	b	c	d	e	f	←Classified as
8	0	0	0	0	0	| a=happy
0	7	0	0	1	0	| b=anger
1	0	7	0	0	0	| c=sad
1	0	0	7	0	0	| d=disgust
1	0	0	0	7	0	| e=fear
1	0	0	0	0	7	| f=surprise

Table 2: Confusion matrix for facial expression classifier.

4.3 Validating Energy Data

We used the mean of the energy data for different affective states as the second feature for our Naïve Bayes fusion. We found an interesting relationship between the energy and the different categories. Figure 3(a) shows the energy mean for different annotations by different PAs. Each point represents a particular annotation by any PA. It was difficult to visually distinguish the energy for the different categories.



(a) (b)

Figure 3: Accelerometer Energy for Different Affective State: (a) Accelerometer energy for six basic emotions. (b) Mean of energy for different affective state.

However, for the three categories, namely happy, sad and anger; we found an important relationship. The mean of energy of sad is much lower than the mean of the energy of happy and that of anger. On the other hand, the mean of energy of anger is not very high where in Russell's two dimensional space it is considered higher than the happy state. We conclude, from our data that happiness usually has a high energy expenditure relative to sadness, which is in line with Russell's theory.

This relationship is best shown in Figure 3(b), where we plot only the mean for different affective states. The horizontal axis represents different emotional states and the vertical axis represents the energy mean for the corresponding emotion. We find that sadness has much lower mean of energy than that of happiness. Also, fear has high value in arousal space and we found that the mean to be much higher than happy and sad. This is also in line with the

Russell's circumplex model where fear is phrased as afraid [Figure 1].

4.4 Performance of Multimodal System

The last part of the discussion addresses the performance of the multimodal system. We see how our system performs with the Naïve Bayes fusion. We find that the system performance for correctly classifying the instances for our training database increases from 89% to 93%. Out of 48 total instances, 45 were classified correctly.

A close analysis from the confusion matrix of the multimodal system from Table 3 gives us the reason of the improvement.

First, the image previously misclassified as fear instead of anger is classified correctly now. The reason is that the mean of energy for anger (8.56E+00) is much lower than that of fear (4.12E+01). As a result, even if the image was not clear enough, it is correctly classified. The same reasoning is also true for the data that were previously misclassified as anger instead of fear.

This data is now also classified correctly. Another interesting observation is the confusion matrix entry for disgust. One disgust entry was misclassified in the unimodal system as anger.

a	**b**	**c**	**d**	**e**	**f**	**←Classified as**
8	0	0	0	0	0	| a=happy
0	8	0	0	0	0	| b=anger
1	0	7	0	0	0	| c=sad
0	0	0	7	0	1	| d=disgust
0	0	0	0	8	0	| e=fear
0	0	0	1	0	7	| f=surprise

Table 3: Confusion matrix using Naïve Bayes classifier.

In the multimodal system it still is misclassified, but to a different class, fear. We observe that the mean of energy for disgust is equidistant from both anger and fear. As a result, the system could not find a close match for this annotation.

5. FINDINGS AND DISCUSSION

5.1 Inherent Theory of Emotion is Not Established

The theory of emotion is not established yet. Psychologists have different approaches to identify different emotions. Research in the field of affective computing is about finding the features that are most likely related to emotion-oriented computing. Understanding those ideas and adapting those to any computational methods is still in progress. Furthermore, expression of emotion greatly varies from person to person, man and women, and also among different age groups and races.

Paul Ekman has identified six basic emotions for psychologists to identify from video sequence using Facial Action Coding System (FACS). Those are happiness, sadness, disgust, anger, fear and surprise. There are other emotions important for automatic detection of emotions like boredom, frustration, excitement and many more. Even Ekman expanded his list of basic emotions to include other emotions like amusement, contempt, embarrassment, excitement, guilt, satisfaction etc.

There are also different approaches in computing for different theories in psychology. For Ekman's FACS to be implemented; feature extraction is needed from facial image and then it needs to be classified. However, there are different emotions with overlapping Coding Schemes which makes the implementation complicated.

For the holistic approach different machine learning algorithms are used. We have used such an approach.

5.2 Multimodal System Needs More Modalities

In our approach, we have argued that multimodal emotion recognition will contribute to the more accurate affective classification. For that we might have to put different weights for different modalities. Also, in person to person communication, we may or may not look for the same features in multiple channels like facial expression, speech and body movements. More importance might be needed for finding same emotional cues in multiple modalities. Again, this varies a lot among person to person. People tend to understand about others emotion from facial expression, tone, body movement, gestures and most importantly context. Depending upon context, the interpretation of a message could be quite different from another. A combination of low level features, high level reasoning, and natural language processing is likely to provide best multimodal affect recognition. But very few systems have been developed in a natural environment considering multiple modalities. Even if they were developed, their performance is measured in a laboratory environment, which might be quite different than in a natural environment.

5.3 Privacy

We have argued that affect detection is important but that also comes with increasing concerns about privacy awareness of the people. However, this argument can be contrasted with the fact that in our system, detected affective state is shared only by the permission of the user. Nevertheless, there remains significant scope for research regarding privacy issues and different levels of anonymization techniques to be dealt with.

6. CONCLUSION

We provided much attention to validate the 'ground truth' data, we found that some emotional states are ambiguous and even human can not identify the emotional states properly. This is because human might have mixed emotions at a particular time. There are no borders with different affective states. However, still we emphasized on the labeling of the emotion by the PAs. Depending upon the interview with the researcher, we incorporated that data for our training database or not. The success of the system largely depends on the emotional self-awareness of the PAs. We think that instead of a particular classification of a particular affective state, one particular instance should be labeled with the different probabilities of falling into different categories. Depending upon those probabilities, machine interpretation of affective states can be applied to human computer interaction. Also, affect sensitive applications should be developed targeting the application scenario. For example, the application for advertisement in smart phones may not be feasible for detecting boredom in a learning environment. We also find that arousal can be captured easier than valence. One such application might be capturing arousal from pupil size which is also a good approximation of the arousal space in psychology. However, for mobile devices it might not be appropriate because of the change of lighting and all other conditions. We believe that by real time sensing of affective states using smart phones, we can machine interpret human affective states and machines can understand part of larger human intelligence. With the continuous advancement of sensor technologies in smart phones, we can predict human affective states more accurately and the application of such affect detection technique might be huge.

7. REFERENCES

[1] Bradley, M. M., Miccoli, L., Escrig, M. A., Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. Center for the Study of Emotion and Attention, University of Florida, 2008.

[2] Fujiki, Y. iPhone as a Physical Activity Measurement Platform. In *CHI 2010(Student Research Competition)*, 2010.

[3] Hansen, F., Christensen, S. R. Emotions, Advertising and Consumer Choice. Copenhagen Business School Press, 2007.

[4] Healey, J. Recording Affect in The Field: Towards Methods and Metrics for Improving Ground Truth Labels. In *Proc. of the Affective computing and intelligent interactions*, 2011.

[5] Healey, J., Nachman, L., Subramanian, S., Shahabdeen, J., Morris, M. Out of the Lab and into the Fray: Towards Modeling Emotion in Everyday Life. In *Proc. 8th International Conference on Pervasive Computing*, 2010.

[6] James, A., Sebe, N. Multimodal Human Computer Interaction: A Survey. In *Computer Vision and Image Understanding*, volume 108, October 2007.

[7] Kapoor, A., Picard, R. Multimodal Affect recognition in learning environments. pages 677–682, 2005.

[8] Kording, K.P., Wolpert, D.M. Bayesian decision theory in sensorimotor control. In *Trends in Cognitive Sciences*, pages 319–326, 2006.

[9] Metallinou, A., Narayanan, S., Lee, S. Decision Level Combination of Multiple Modalities for recognition and analysis of emotional expression. In *Proc. of the International Conference on Acoustics, Speech, and Signal Processing*, 2010.

[10] Monwar, M., Prkachin, K., Rezaei, S. Eigenimage Based Pain Expression Recognition. In *International Journal of Applied Mathematics*, May 2007.

[11] Nicolaou, A., Pantic, M., Gunes, H. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. In *IEEE Transactions On Affective Computing*, volume 2, 2011.

[12] Pantic, M., Roisman, G., Huang, T., Zeng, Z. A survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 31, January 2009.

[13] Picard, R. W. Affective Computing. The MIT Press, 1997.

[14] Picard, R. W. Affective Computing: Challenges. In *International Journal of Human-Computer Studies*, 2003.

[15] Rabbi, M., Ali, S., Choudhury, T., Berke, E. Passive and In-situ Assessment of Mental and Physical Well-being using Mobile Sensors. In *Proc. of the 13th international conference on Ubiquitous computing*, 2011.

[16] Rime, B., Mesquita, B., Philipot, P. Long lasting cognitive and social consequences of emotion: Social sharing and Rumination. In *European Review of Social Psychology*, 1992.

[17] Rime, B., Philippot, P., Zech, E., Luminet, O., Finkenauer, C. Social sharing of emotion: New evidence and new questions. In *European review of social psychology*, volume 9, pages 145–189.

[18] Russell, J. A. A circumplex model of affect. In *Journal of Personality and Social Psychology*, volume 39, 1980.

[19] Scherer, K.R., Ekman, P. Methods For Measuring Facial Action. In *Handbook of methods in nonverval behavior research*, pages 45–135. Cambridge University Press, 1982.

[20] Serre, T., Bouvrie, J., Ivanov, Y. . Error weighted Classifier Combination for Multimodal Human Identification. In *Tech. Rep., MIT, Cambridge, MA*, 2005.

[21] Sharma, R., Huang, T., Pavlovic, V. Toward Multimodal Human Computer Interface. In *Proc. IEEE*, volume 86, pages 853–869, 1998.

[22] Tian, Y., Cohn, J., Kanade, T. Facial Expression Analysis. In *Handbook of Face Recognition*. Springer, 2005.

[23] Tian, Y., Cohn, J., Kanade, T. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. In *IEEE Transactions on Affective Computing*, volume 1, 2010.

[24] A. E. K. N. Wagner, J. From Physiological Signals to Emotions Implementing and Comparing Selected Methods for Feature Extraction and Classification. In *IEEE Int'l Conf. Multimedia and Expo*, pages 940–943, 2005.

[25] M. Weiser. Some computer science issues in ubiquitous computing. In *Communications of the ACM*, pages 74–83, 1993.