IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 23-30
www.iosrjournals.org

Improved Affect Detection Algorithm On Face Patches Using Local Binary Patterns and PCA

[1]Dhananjay Theckedath, [2]Dr. R.R Sedamkar
[1]PhD Scholar with Thakur College of Engineering
[2]Professor with Thakur College of Engineering and is also the Dean-Academic

Abstract: *Detection of learning centered affect is an important element in developing intelligent educational interfaces that are capable of responding to the affective needs of students. There exists an inseparable link between emotions and cognition and the field of Affective Computing (AC) aspires to narrow the communicative gap between human beings and computers. This is achieved by developing computational systems that recognize and respond to the affective states of the user. Learning centered affective states include Boredom, Confusion, Delight, Engagement and Frustration. This paper uses the Local binary patterns (LBP) ans a feature extraction algorithm and PCA for dimentionality reduction to detect confusion. This algorithm is applied to the full face as well as Lip and eye sections of the face. The classifier that is used is the chi-square classifier. The template matching achieved the generalization performance of 81.5% for 2-class task and 76.2 % for the 4-class on the full face. Template matching on the Lip patch achieved the generalization performance of 79 % for 2-class task which involved Confused and not confused The results obtained suggest that instead of working with the entire face, we could work only with the lip portion to identify the affect states as it would computationally be more efficient.*

I. Introduction

One of the areas of Human Computer Interface (HCI) in which progress is being made is in a virtual classroom environment where learning typically means using a computer to deliver part, or all of a course [1]. In such a scenario, it becomes imperative to evaluate the extent to which the student has learnt and also find out ways to improve the teaching learning process. Affect detection which is centered around learning is a major component in developing educational interfaces that are capable of responding to the learning needs of students [2].

Predicting learning centered affect states while students interact with the system is a challenging problem. While there has been some work carried out in detecting six basic emotions which are Happy, Sad, Surprise, Anger, Fear and Disgust , there is very little research reported on detection and classification of learning centered affective states [3]. Affect states pertaining to learning are known to be different than the standard basic emotions. Learning centered affective states include Boredom, Confusion, Delight, Engagement and Frustration. While basic emotions can be detected fairly accurately, it is still unclear if learning centered affective states can be detected with the same fidelity. For basic emotions the links between emotion and expression have been carefully mapped. Similar mapping for learning centered affective states is largely missing and is an open question. Due to the large imbalances in the distribution of affect, accuracy for learning centered affect detectors also varies widely.

There has been a recent development of using Convolutional Neural Networks in facial expression detection. Classification task are being explored using Deep learning and the results have been encouraging [4]. With the evolution of deep learning in computer vision, emotion recognition has become a widely-tackled research problem [5].

II. Expression Analysis Considered For The Hci Point Of View

The expression of emotion is achieved through a complex combination of information produced from the body and the brain. However the scientific community has yet to measure and quantify the relationship between facial expressions and emotions. There have been several attempts to quantify this relationship. One such method uses objective coding schemes based on visible units of facial behavior. This is usually done by analyzing videos, frame-by-frame which are recorded during an experiment. Each frame is analyzed in a detailed manner by human coders who systematically follow descriptive rules of judgment.

Systematic measurements were created by Ekman and Friesen [6], which has proved to be the bench mark for subsequent studies in Expression analysis. They introduced a new system known as the Facial Action Coding System (FACS). The Facial Action Coding System proposed by them partitions the visible effects of facial muscle activation into "Action units" (AU). In this system, each action unit is related to one or more facial muscles. The Facial Action Coding System (FACS) is a broad, anatomically based system which is used for measuring nearly all visually noticeable facial movements. The latest FACS system describes facial activity based on 44 unique Action Units.

While there has been some work carried out in detecting six basic emotions which are Happy, Sad, Surprise, Anger, Fear and Disgust , there is very little research reported on detection and classification of learning centered affective states. Affect states pertaining to learning are known to be different than the standard basic emotions. Learning centered affective states include Boredom, Confusion, Delight, Engagement and Frustration.[7]

It is a matter of further investigation whether these states can be detected with the same accuracy as the basic emotions, where the association between emotion and expression have been precisely mapped. Similar mapping for learning centered affective states is largely missing and is an open question. Due to the highly skewed class distribution of learning centered affective states, designing classifiers is a challenging task and the accuracy for detecting learning centered affect also varies widely [8].

III. Relationship Between Action Units And Emotions

Correlations were computed to determine the extent to which each of the Action Units were diagnostic of the affective states of boredom, confusion, delight, frustration, and neutral.

In the paper [9].Two sets of correlations were computed in order to determine whether significant patterns emerged across both independent coders. The analysis revealed that there was a good degree of concordance among the two judges. The paper mentions that presence of AU 7 (lip tightener), AU 12 (lip corner puller), AU 25 (lips part), and AU 26 (jaw drop) coupled with an absence of AU 45 (blink) segregate delight emotion from neutral. Similarly Confusion was manifested by a lowered brow (AU 4), the tightening of the eye lids (AU 7), and a notable lack of a lip corner puller (AU 12). Frustration is a state that is typically associated with significant physiological arousal, yet the facial features tracked were not very good at distinguishing this emotion from neutral. The only significant correlation with frustration was obtained for AU 12 (lip corner puller) – perhaps indicative of a half smile. Boredom is not easily distinguishable from neutral on the basis of the facial features. Boredom which resembles an expressionless face has no single action unit found to be associated with it [10].

Action Unit	**Face association**	**Associated Affect**
AU1	Inner Brow Raiser	Confusion
AU4	Brow Lowerer	Boredom, Confusion, Delight, Neutral
AU7	Lid Tightener	Boredom, Confusion, Delight, Neutral
AU12	Lip Corner Puller	Boredom, Confusion, Delight, Neutral, Frustration
AU25	Lips Part	Delight, Neutral
AU26	Jaw Drop	Boredom, Confusion, Delight, Neutral
AU43	Eye Closure	Frustration, Neutral
AU45	Blink	Boredom, Delight

Table 1: Correlations between action units and affective states

IV. Proposed Methodology

This paper aims at detecting learning centered affect state viz. confusion. We have used Local binary patterns on the full face as well as face patches particularly lip and eye portions of the image. The local binary pattern operator is an image operator which transforms an image into an array or image of integer labels describing small-scale appearance of the image. These labels or their statistics, most commonly the histogram, are then used for further image analysis. The most widely used versions of the operator are designed for monochrome still images but it has been extended also for colour (multichannel) images as well as videos and volumetric data.

The basic local binary pattern operator [11], was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. The LBP operator can be used to detect micro-pattern primitives like the ones shown in Fig 1.

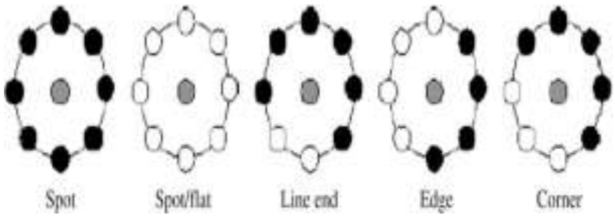


Fig.1. LBP Micropatterns

The original version of the local binary pattern operator works in a 3×3 pixel block of an image. The pixels in this block are thresholded by its centre pixel value, multiplied by powers of two and then summed to obtain a label for the centre pixel. As the neighbourhood consists of 8 pixels, a total of 28 = 256 different labels can be obtained depending on the relative gray values of the center and the pixels in the neighbourhood. The LBP operator labels each pixel of an image by thresholding a 3 x 3 neighbourhood of each pixel with the centre value and considering the result as a binary number. Implementing the LBP gives us blocks of constant grey values. This is shown in Fig. 2.

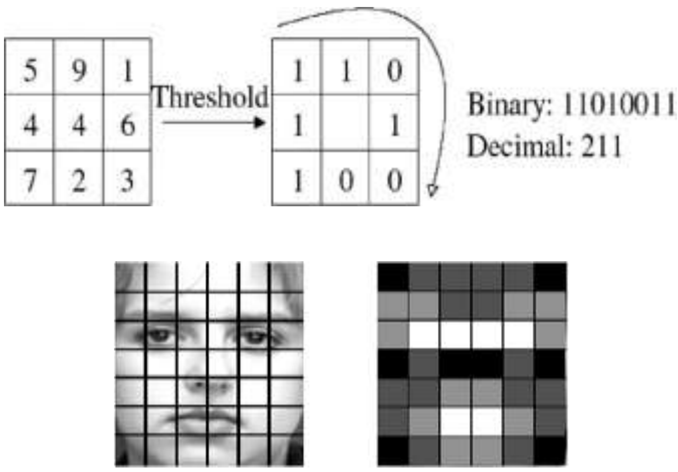


Fig. 2. LBP representation

For a 8 bit image, the LBP operator, produces 28 = 256 different output values. It has been shown that certain bits contain more information than others. Therefore, it is possible to use only a subset of the 28Local Binary Patterns to describe the texture of images. These fundamental patterns are known as uniform patterns. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. It is observed that uniform patterns account for nearly90% of all patterns in the (8, 1) neighborhood and for about 70%in the (16, 2) neighborhood in texture images. LBP features extracted from each sub-region are concatenated into a single, spatially enhanced feature histogram. This LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image and represents the local texture and global shape of face images.

For a 8 bit image, each feature vector is basically a histogram obtained from the LBP image and hence is of size 1 x 256. This results in a very long feature vector. We used the Principle component analysis for dimensionality reduction. We computed the PCA of the histogram values and selected only the first 10 values which formed the new feature vector. We did this for the full face as wel as the eyes and Lip portion of the image [12]

Fig. 3. Block diagram of LBP+PCA

4.1.1 Feature Extractin Using Chi Square Method

Template matching is a technique in digital image processing for finding small parts of an image which match a template image. Template matching was used in this paper to perform face recognition using the LBP-based facial representation. Fig 3. is the block diagram of the entire process. A template is formed for each class of face images and a nearest-neighbour classifier is used to match the input image with the closest template. Here we first adopted template matching to classify facial expressions for its simplicity. In training, the histograms of expression images in a given class were averaged to generate a template for this class.
Following this , the Chi square statistic as the dissimilarity measure for histograms was selected:

$$\chi 2\ (S,M)= \sum_{x=0}^{N-1} \frac{(S_{ij} - M_{ij})^2}{(S_{ij} + M_{ij})} \qquad (1)$$

where S and M are two LBP histograms.
Given below is the flow chart of the algorithm used.

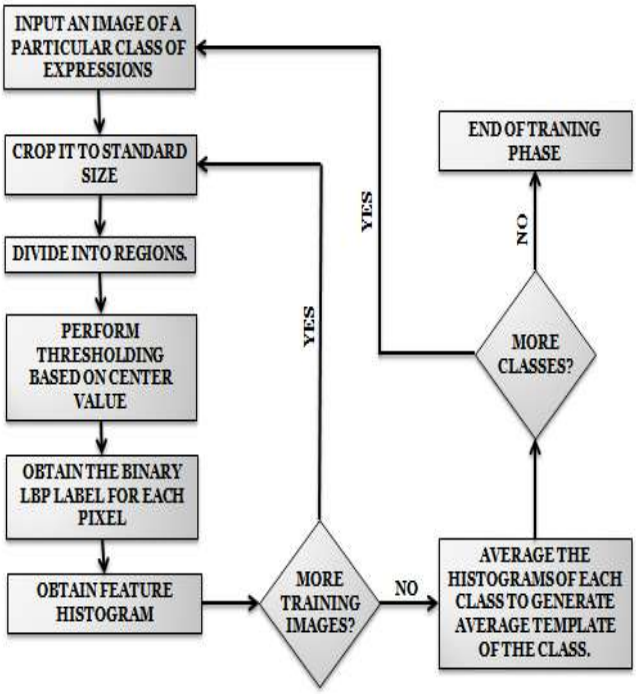


Fig.4. Flowchart of the algorithm- Training

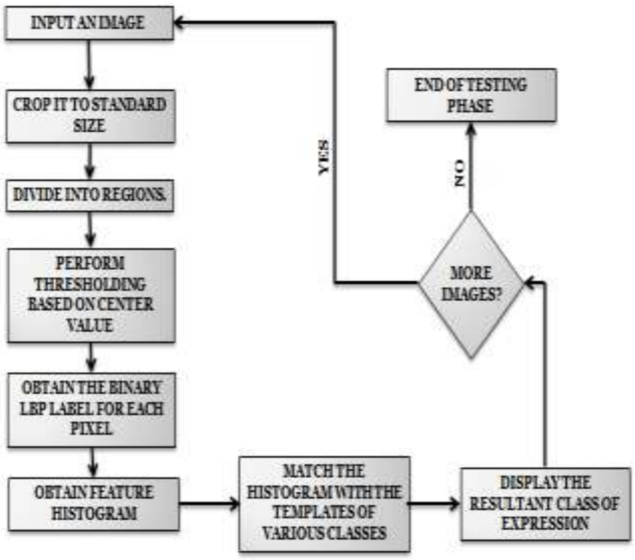


Fig.5. Flowchart of the algorithm- Testing

V. Result Analysis And Dicussion

Images are acquired from the Cohn-Kanade database [13] which consist of 87 distinct images with different expressions. All images are in PNG format. And the size of each image is 640×490, with 256 grey levels. The data base was shown to four professors and four students and they were asked to select images that they felt represented confusion. Only those in which there was to complete unanimity were chosen to represent confusion [14]. The database consists of face images against a back ground. Initial pre-processing was carried out. This involves separating the face region and then clipping the eyes and the lip portions as these are the two regions from which affect detection can be carried out. The Viola Jones algorithm [15] is used for separating the face as well as the eyes and the lips portion Fig.6.

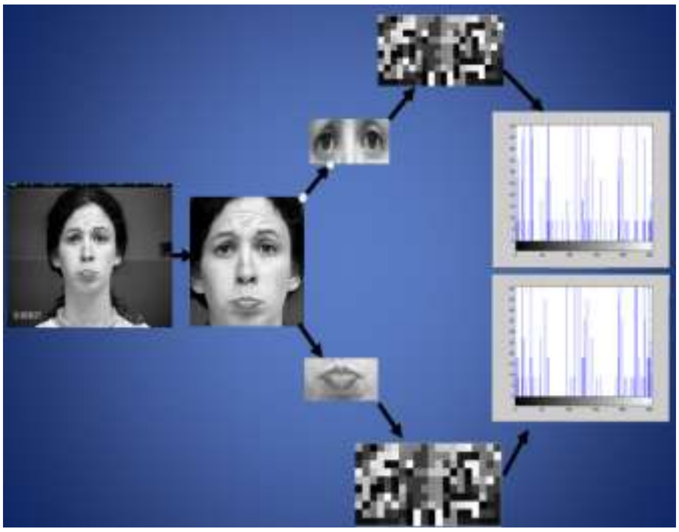
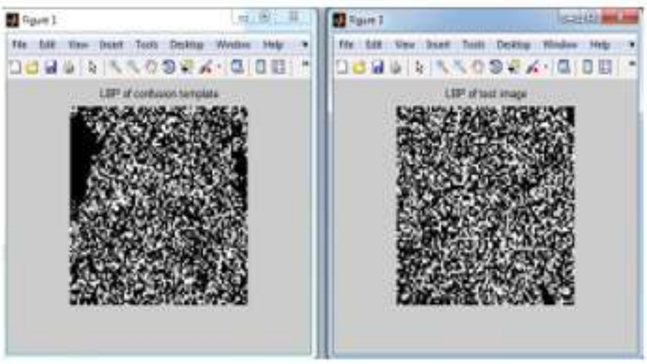


Fig.6. Result of the Algorithm

The full face LBP was implemented on the full face and the histogram e was extracted , Fig.7 and the LBP image was obtained. Refer Fig. 8 and Fig. 9. The Chi square statistic as the dissimilarity measure was used on the obtained histograms.

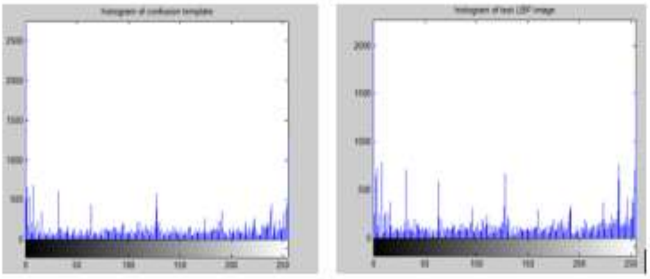
(a) (b) (c) (d)

Fig. 7. Frontal face of various expressions (a) Happy (b) Sad (c) Shock (d) Confused



(a) (b)

Fig. 8. LBP image: (a) LBP of the confusion template, (b) LBP of test image



(a) (b)

Fig. 9. LBP Histogram; (a) Confusion template, (b) test image

The Algorithm was tested on a 2- Class dataset comprising of Confusion and Happy as well as a 4 - Class data set comprising of Confusing, Happy, Surprise and Sad. The template matching achieved the generalization performance of 81.5% for 2-class task, refer Table 3. The performance of the algorithm fell to 76.2 % for the 4-class. It was observed that the Sad class brought down the performance. The confusion matrix created is shown in Table 3.

Methods (feature + Classifier)	2-Class Recognition (%)	4 –Class recognition (%)
LBP + Template Matching	81.5	76.2

Table 2. Comparison of results based on 2 class recognition and 4 class recognition

	Confusion (%)	Happy (%)	Surprise (%)	Sad (%)
Confusion(%)	83.4	1.1	0	15.5
Happy(%)	1.5	91.5	6.5	0.5
Surprise(%)	2	3.5	94.5	0
Sad(%)	19.5	2.5	2	76.1

Table 3. Confusion matrix of 4 class facial expression

The same algorithm was then implemented on only the Lip patch of the face. In this, the Lip portion was cropped form the face using the Viola Jones algorithm, Fig. 10.

(a) (b) (c) (d)

Fig. 10. (a) Lip portion,(b) LBP of confusion Lip template (c) Test image (d) LBP of test image

The template matching algorithm on the Lip portion achieved the generalization performance of 79 % for 2-class task which involved Confused and not confused.

Methods (feature + Classifier)	2-Class Recognition (%)	4 –Class recognition (%)
LBP + Template Matching	79	74.1

Table 4 Comparison of results based on 2 class recognition and 4 class recognition

We also computed the speed of implementing the LBP+PCA on full face and the Lip portions. While there was no significant deterioration in the results of the face and lips, It was observed that there was a substantial reduction in the processing time when working only with the Lip portion. Fig 11 and Fig 12 give us the time taken for running the algorithm. We can as of now conclude that instead of working with the entire face, it would be prudent to work only with the lip portion to identify the affect states.

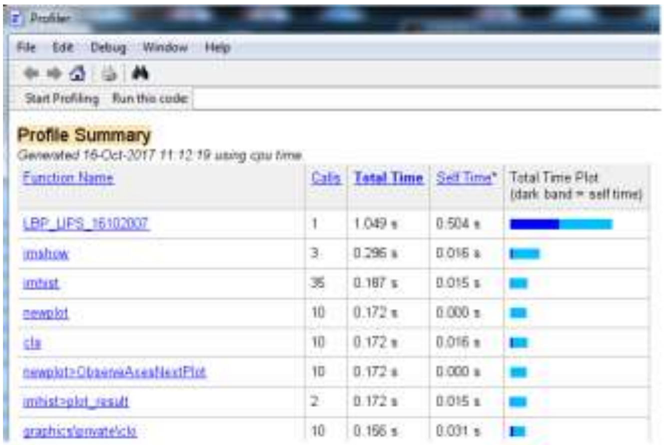


Fig. 11. Total time required to execute LBP on Lip portion

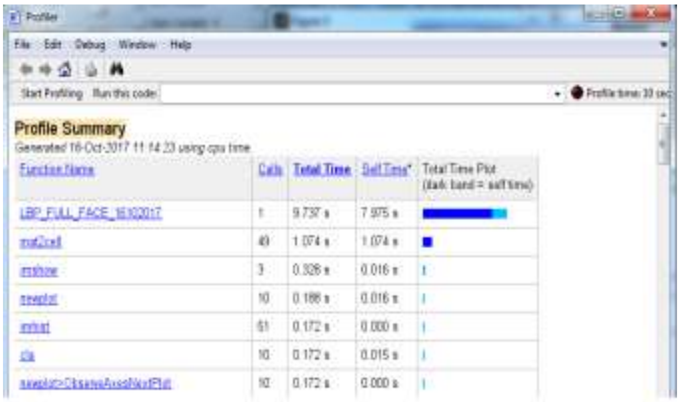


Fig.12. Total time required to execute LBP on Full face

VI. Future Work

We plan to use various hybrid algorithms to detect learning centerd affect states. There is a recent interest toward using Convolution Neural networks and Deep learning. These will be tried out and would be analysed. This research would lead to development of robust Virtual classroom environment in which a computer is used to deliver part, or all of a course whether it is in a school, college, part of mandatory business training or a full distance learning course [16]. This research will help the Industry and also the society as a whole as the cost of education in remote areas will substantially go down.

VII. Conclusion

There has been a growing interest in recent years in improving the interaction between humans and computers. To achieve an effective human-computer intelligent interaction, there is a need for the computer to be able to interact with the user naturally, similar to the way human-human interaction takes place. The paper describes the LBP algorithm couples with PCA to identify confusion. It has been observed that applying the algorithm only he Lip portion also gives us favourable results which can be further improved upon by using other hybrid techniques and deep learning algorithms on only sections of the entire face.

References

[1] B.Wolf, Burleson, Arroyo, Dragon, Cooper and Picard. "Affect aware tutors: Recognising and responding to student affect".International journal of Learning technology,4(3/4):129-164, 2009

[2] Gong Chao ; Beijing Inst. of Technol., Beijing, China. "Human-computer interaction: The usability test methods and design principles in the human-computer interface design." Aug. 2009 283 – 285.

[3] Sangeeta Kakoty, Monohar Lal, Shikhar Kr. Sarma "E-learning as a Research Area: An Analytical Approach", International Journal of Advanced Computer Science and Applications, Vol. 2, No. 9, 2011.

[4] Yichuan Tang," Deep learning using Linear support Vector Machines, Cornell University Library, arXiv:1306.0239v4

[5] Prudhvi Raj Dachapally, "Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units", Cornell University Library arXiv:1706.01509

[6] Ekman, P., Friesen, W.: "Facial action coding system: A technique for the measurement of facial movement". Consulting Psychologists Press, Palo Alto (1978)

[7] Bethany McDaniel, Sidney D'Mello, Brandon King, Patrick Chipman, Kristy Tapp, Art Graesser, "Facial Features for Affective State Detection in Learning Environments", Proceedings of the Annual Meeting of the Cognitive Science Society 2007.

[8] Brandon King, Patrick Chipman, Kristy Tapp, Art Graesser. "Facial Features for Affective State Detection in Learning Environments", Journal Issue: Proceedings of the Annual Meeting of the Cognitive Science Society, 29. 2007

[9] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang,"A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", IEEE Transactions On Pattern Analysis And Machine Intelligence, VOL. 31, NO. 1, Jan 2009

[10] Michel F. Valstar, Maja Pantic, "Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database", Proc. 3rd Intern. Workshop on EMOTION 2010.

[11] M.Topi, O.Timo, P.Matti, "Robust texture classification by subsets of local binary patterns", Pattern Recognition, 2000. Proceedings. 15th International Conference , 2000.

[12] S L Happy,Aurobinda Routray, "Automatic Facial Expression Recognition Using Features of Salient Facial Patches", IEEE Transactions On Affective Computing, Vol. 6, No. 1, January-March 2015.

[13] P. Lucy, J.F.Cohn, T. Kanade "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression "Computer Vision and Pattern Recognition Workshops (CVPRW), 2010.\

[14] Chih-Hung Wu, Yueh-Min Huang and Jan-Pan Hwang, "Review of affective computing in education/learning: Trends and challenges" British Journal of Educational Technology (2015) doi:10.1111

[15] Paul Viola, Michael Jones "Rapid Object Detection using a Boosted Cascade of Simple Features" Conference On Computer Vision And Pattern Recognition 2001.

[16] Ramkumar, Sridhar Iyer et all. "A theory driven approach to predict frustration in an ITS" , IEEE transactions on learning technologies; 6, 4; 378-388; 2013