Emotion Detection using Sobel Filtering and Retrieving with Sparse Code words

Mahevish Fatima Mohammed Gani (Student)
Computer Science and Engineering, Maharashtra Institute of Technology, Aurangabad, Maharashtra
Email. - mahvishfatema@gmail.com

Abstract: Extracting and understanding of emotion is of high importance for the interaction among human and machine communication systems. The most expressive way to display the human's emotion is through facial expression analysis. This paper presents and implements an automatic extraction of facial expression and emotion from still image. There are steps to detect the facial emotion; (1) Preprocessing, skin colour segmentation and edge detection using Sobel filtering and (2) Verifying the emotion of characteristic with Bezier curve. Retriving emotion using sparse code words which are used to analyze this results in medicine, and enhancement in CBIR.
To evaluate the performance of the proposed algorithm, we assess the ratio of success with emotionally expressive facial image database. Experimental results shows average 66.6% of success to analyze emotion.
Key Words: Facial Expression, Preprocessing ,Skin Colour Segmentation, Sobel Filters, Bezier Curve, Attribute enhanced sparse code word.

1. INTRODUCTION:

Emotion plays an important role in human communication and therefore also human machine dialog systems can benefit from affective processing and also in communication media. Faces are the most relevant social stimuli as they communicate information essential for the course of social interaction and communication. Everyday almost everyone in this world interact with other in one or another way either directly (for e.g. face to face) or indirectly (for e.g. phone calls). In some profession interaction with people are the main deed to perform like call centers, sale executives etc. With great advancement in technology in terms of different techniques of people interacting with each other it is quite necessary that one should be aware of current emotions of the person he/she is interacting. With the advancement of 3G technology in mobile communication field one may be capable to interact face to face with other while talking so if one is aware of mood of other in advance that interaction will certainly result in social as well as professional benefits. Firstly, what are emotions? A mental state that arises spontaneously rather than through conscious effort and is often accompanied by physiological changes and these physiological changes are recognized from outer world. Emotion lies behind much of the richness of human life and is one of the main drivers behind our choices and decisions.

Facial expression convey information about what emotion is currently experienced by a target which affects how the target is percieved and what behavioral tendencies are elicited in the observer. There are many ways that humans can express their emotions. The most natural way to express emotions is using facial expressions. A human can express his/her emotion through lip and eye. Detecting the emotion and retrieve from database using the attribute enhance sparse codeword. It is an enabling technology for many applications including automatic face annotation, E-learning, Tutoring System and Monitoring etc. In human computer interfaces valuable communication media provide natural interaction to the user. Human computer interaction could significantly improved if computer could detect emotions of the users from their facial expression and react according to the users need.

Our main objective, to achieve Emotion detection and large-scale content-based face image retrieval. To compensate this problem ,given a query face image, it detects emotion using sobel filter and Bezier curve after that content-based face image retrieval tries to find similar face images from a large image database. It is an enabling technology for many applications including automatic face annotation etc. To improve content based image retrieval we use attribute enhanced sparse codewords to retrieve similar images from the database. To evaluate the performance of these methods, we conduct extensive experiments on separate public datasets named LFW. These datasets contain faces taken in unconstrained environment.

To sum up the contribution of this work, First, it takes an image, then by skin colour segmentation, it detects human skin colour, then it detect human face. Then it separates the eyes & lip from the face. Then it draws bezier curve for eyes & lips. Then it compares the bezier curve of eyes and lips to the bezier curves of eyes & lips that are stored in the data base. Then it finds the nearest bezier curse from the data base & gives that data base stored bezier curve emotion as this image emotion.

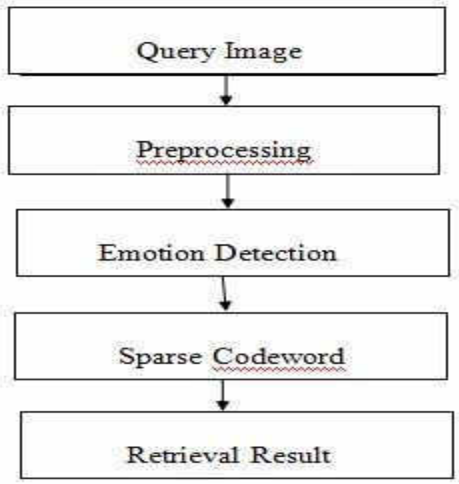

Figure 1: System Overview

2. RELATED WORK:

Facial expressions give important clues about emotions. Therefore, several approaches have been proposed to classify human affective states. The features used are typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio, which use global statistics of the acoustic features. For a complete review of recent emotion recognition systems based on facial expression the readers are referred to [1]. Mase proposed an emotion recognition system that uses the major directions of specific facial muscles [2]. With 11 windows manually located in the face, the muscle movements were extracted by the use of optical flow. For classification, K-nearest neighbor rule was used, with an accuracy of 80% with four emotions: happiness, anger, disgust and surprise. Yacoob et al. proposed a similar method [3]. Instead of using facial muscle actions, they built a dictionary to convert motions associated with edge of the mouth, eyes and eyebrows, into a linguistic, per- frame, mid-level representation. They classified the six basic emotions by the used of a rule-based system with 88% of accuracy. Black et al. used parametric models to extract the shape and movements of the mouse, eye and eyebrows [4]. They also built a mid- and high-level representation of facial actions by using a similar approach employed in [3], with 89% of accuracy. Tian et al. attempted to recognize Actions Units (AU), developed by Ekman and Friesen in 1978 [5], using permanent and transient facial features such as lip, nasolabial furrow and wrinkles [6]. Geometrical models were used to locate the shapes and appearances of these features. They achieved a 96% of accuracy. Essa et al. developed a system that quantified facial movements based on parametric models of independent facial muscle groups [7]. They modeled the face by the use of an optical flow method coupled with geometric, physical and motion-based dynamic models. They generated spatial-temporal templates that were used for emotion recognition. Without considering sadness that was not included in their work, a recognition accuracy rate of 98% was achieved. A method that extracts region of eye and lip of facial image by genetic algorithm has been suggested recently [8]. The obtained results show that the success rate and running speed in face emotion detection using eye and lip by Sobel filtering algorithm and Bezier curve method in comparison with the genetic algorithm has better performance.

3. THE PROPOSED METHOD:

The main goal of this paper is to design a method with a sobel filtering algorithm to emotion detection.

a. Preprocessing

i. Skin Colour Segmentation

For skin colour segmentation, first we contrast the image. Then we perform skin colour segmentation. Then, we have to find the largest connected region. Then we have to check the probability to become a face of the largest connected region. If the largest connected region has the probability to become a face, then it will open a new form with the largest connected region. If the largest connected regions height & width is larger or equal than 50 and the ratio of height/width is between 1 to 2, then it may be face.

ii. Face Detection

For face detection, first we convert binary image from RGB image. For converting binary image, we calculate the average value of RGB for each pixel and if the average value is below than 110, we replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from RGB image. Then, we try to find the forehead from the binary image. We start scan from the middle of the image, then want to find a continuous white pixels after a continuous black pixel. Then we want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then we break the scan because if we reach the eyebrow then this situation will arise. Then we cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width.

In the figure, X will be equal to the maximum width of the forehead. Then we will have an image which will contain only eyes, nose and lip. Then we will cut the RGB image according to the binary image.

iii. Eyes Detection

For eyes detection, we convert the RGB face to the binary face. Now, we consider the face width by W. We scan from the W/4 to (W-W/4) to find the middle position of the two eyes. The highest white continuous pixel along the height between the ranges is the middle position of the two eyes.

Then we find the starting high or upper position of the two eyebrows by searching vertical. For left eye, we search w/8 to mid and for right eye we search mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes. There may be some white pixels between the eyebrow and the eye. To make the eyebrow and eye connected, we place some continuous black pixels vertically from eyebrow to the eye. For left eye, the vertical black pixel-lines are placed in between mid/2 to mid/4 and for right eye the lines are in between mid+(w-mid)/ 4 to mid+3*(w-mid)/ 4 and height of the black pixel-lines are from the eyebrow starting height to (h- eyebrow starting position)/4. Here w is the width of the image and mid is the middle position of the two eyes and h is the height of the image. Then we find the lower position of the two eyes by searching black pixel vertically.

For left eye, we search from the mid/4 to mid - mid/4 width. And for right eye, we search mid + (w-mid)/ 4 to mid+3*(w- mid)/ 4 width from image lower end to starting position of the eyebrow. Then we find the right side of the left eye by searching black pixel horizontally from the mid position to the starting position of black pixels in between the upper position and lower position of the left eye. And left side for right eye we search mid to the starting position of black pixels in between the upper position and lower position of right eye. The left side of the left eye is the starting width of the image and the right side of the right eye is the ending width of the image. Then we cut the upper position, lower position, left side and the right side of the two eyes from the RGB image.

Eye display a strong vertical edges (Horizontal transitions) and white part of eye.The Sobel mask can be applied to an image and the horizontal projection of vertical edges can be obtained to determine the Y coordinate of the eyes. Sobel edge detection is applied to the upper half of the face image and the sum of each row is horizontally plotted.

-1	0	1
-2	0	2
-1	0	1

Vertical edger

1	2	1
0	0	0
-1	-2	-1

Horizontal edger

Figure 2: The Sobel Mask

The peak with the lower intensity value in horizontal projection of intensity is selected as the Y coordinate.

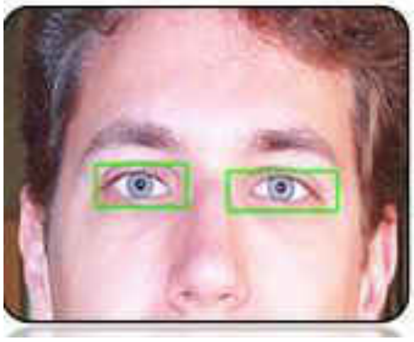

Figure 3 : Eyes Detection

iv. Lip Detection

For lip detection, we determine the lip box. And we consider that lip must be inside the lip box. So, first we determine the distance between the forehead and eyes. Then we add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the ¼ position of the left eye box and ending point will be the ¾ position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then we will cut the RGB image according the box. So, for detection eyes and lip, we only need to convert binary image from RGB image and some searching among the binary image. Sobel edge detection is applied to the lower half of the face image and the sum of each row is horizontally plotted.

Figure 4: Lip Detection

v. Apply Bezier Curve on Lip

In the lip box, there is lip and may be some part of nose. So, around the box there is skin colour or the skin. So, we convert the skin pixel to white pixel and other pixel as black. We also find those pixels which are similar to skin pixels and convert them to white pixel. Here, if two pixels RGB values difference is less than or equal 10, then we called them similar pixel. Here, we use histogram for finding the distance between the lower average RGB value and higher average RGB value. If the distance is less than 70, then we use 7 for finding similar pixel and if the distance is getter than or equal 70 then we use 10 for finding similar pixel. So, the value for finding similar pixel depends on the quality of the image. If the image quality is high, we use 7 for finding similar pixel and if the image quality is low, we use 10.

So, in the binary image, there are black regions on lip, nose and may some other little part which have a little different than skin colour. Then we apply big connected region for finding the black region which contain lip in binary image. And we are sure that the big connected region is the lip because in the lip box, lip is the largest thing which is different than skin.

Then we have to apply Bezier curve on the binary lip. For apply Bezier curve, we find the starting and ending pixel of the lip in horizontal. Then we draw two tangents on upper lip from the starting and ending pixel and also find two points on the tangent which is not the part of the lip. For the lower lip, we find two point similar process of the upper lip. We use Cubic Bezier curves for draw the Bezier curve of the lip. We draw two Bezier curve for the lip, one for upper lip and one for lower lip.

vi. Apply Bezier Curve on Eye

For apply Bezier curve on eyes, first we have to remove eyebrow from eye. For remove eyebrow, we search 1st continuous black pixel then continuous white pixel and then continuous black pixel from the binary image of the eye box. Then we remove the 1st continuous black pixel from the box and then we get the box which only contains the eye.

Now, the eye box which contains only eye, has some skin or skin colour around the box. So, we apply similar skin colour like the lip for finding the region of eye. Then we apply big connect for finding the highest connected region and this is the eye because in the eye box, eye is the biggest thing which is not similar to the skin colour.

Then we apply the Bezier curve on the eye box, similar to the lip. Then we get the shape of the eye.

Vii. Database and Training

In our database, there are two tables. One table "Person" is for storing the name of people and their index of 4 kinds of emotion which are stored in other table "Position". In the "Position" table, for each index, there are 6 control points for lip Bezier curve, 6 control points for left eye Bezier curve, 6 control points for right eye Bezier curve, lip height and width, left eye height and width and right eye height and width. So, by this method, the program learns the emotion of the people.

b. Emotion Detection

For emotion detection of an image, we have to find the Bezier curve of the lip, left eye and right eye. Then we convert each width of the Bezier curve to 100 and height according to its width. If the person's emotion information is available in the database, then the program will match which emotion's height is nearest the current height and the program will give the nearest emotion as output.If the person's emotion information is not available in the database, then the program calculates the average height for each emotion in the database for all people and then get a decision according to the average height.

c. Sparse Coding:

In this section, we first describe how to use sparse coding for face image retrieval. We then describe details of the proposed attribute-enhanced sparse coding. We apply the same procedures to all patches in a single image to find different codewords and combine all these codewords together to represent the image.

i. **Sparse coding for face image retrieval (SC)**

Using sparse coding for face image retrieval, we solve the following optimization problem:

$$\min_{D,V} = \sum_{i=1}^{n} \left|\left|x^{(i)} - Dv^{(i)}\right|\right|2 + \lambda||v(i) \quad \text{subject to } \left|\left|D * j\right|\right|2 = 1, \forall j \qquad (1)$$

where x(i) is the original features extracted from a patch of face image i, D ∈ Rd×K is a to-be-learned dictionary contains K centroids with d dimensions.V = [v(1),v(2),...,v(n)] is the sparse representation of the image patches. The constraint on each column of D (D∗j) is to keep D from becoming arbitrarily large. Using sparse coding, a feature is a linear combination of the column vectors of the dictionary provides an efficient online algorithm for solving the above problem. Equation (1) actually contains two parts: dictionary learning (find D) and sparse feature encoding (find V). Coates et. al. found that using randomly sampled image patches as dictionary can achieve similar performance as that by using learned dictionary (< 2.7% relative improvement in their experiments) if the sampled patches provide a set of overcomplete basis that can represent input data. Because learning dictionary with a large vocabulary is time-consuming (training 175 codebooks with 1600 dimension takes more than two weeks to finish), we can just use randomly sampled image patches as our dictionary and skip the time-consuming dictionary learning step by fixing D in the Equation (1) and directly solve V . When D is fixed, the problem becomes a L1 regularized least square problem, and can be efficiently solved using LARS algorithm .After finding v(i) for each image patch, we consider nonzero entries as codewords of image i and use them for inverted indexing. Note that we apply the above process to 175 different spatial grids separately, so codewords from different grids will never match. Accordingly, we can encode the important spatial information of faces into sparse coding. The choice of K is investigated . We use K = 1600 in the experiments, so the final vocabulary size of the index system will be 175 × 1600 = 280,000.

4. EXPERIMENTAL RESULT:

After applying the operations on real time images and stable images (taken from local drive), this system successfully works on all the modules and detects the all type of emotions (i.e. Smile, Sad, Normal, and Surprise) that are mentioned in the paper.

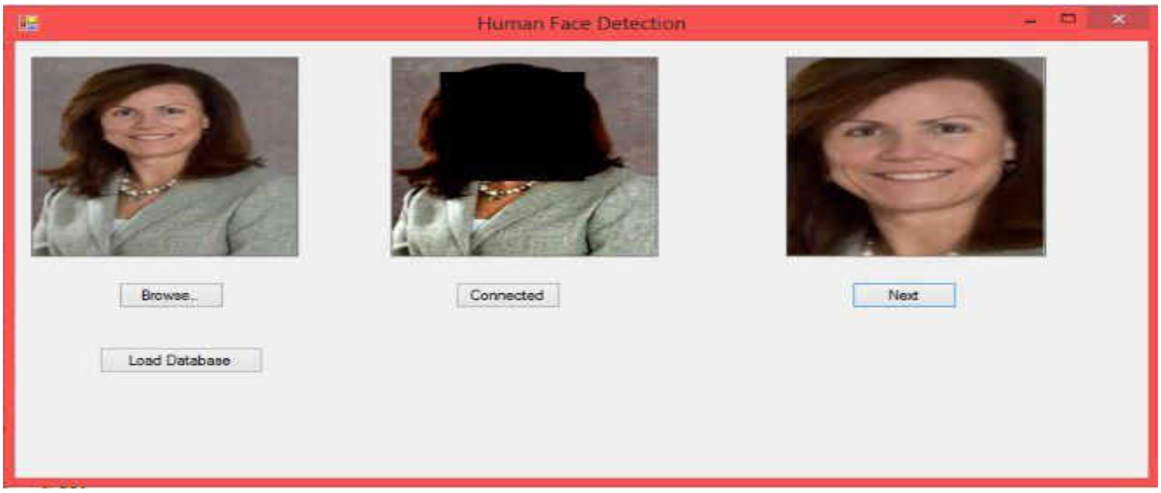

Figure 5: Form 1 for Emotion Detection System

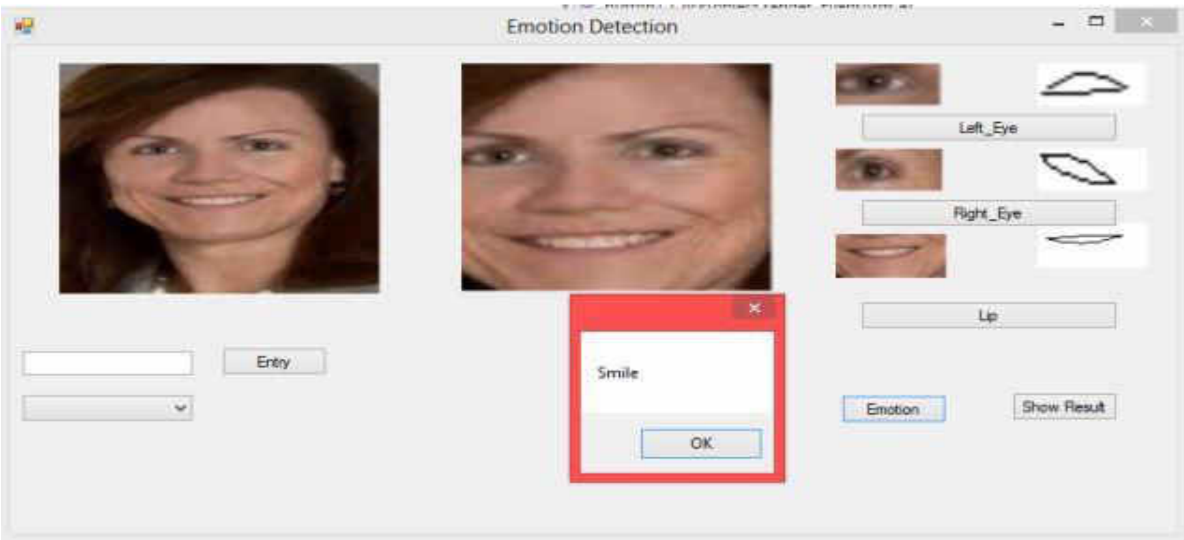

Figure 6: Form 2 for Emotion Detection System

In the above form actual face is detected and then with the help of (eye_lip button) cubic bezier curve we have to found out the exact eye and lip coordinates. After that emotion will be detected and displayed on the screen.

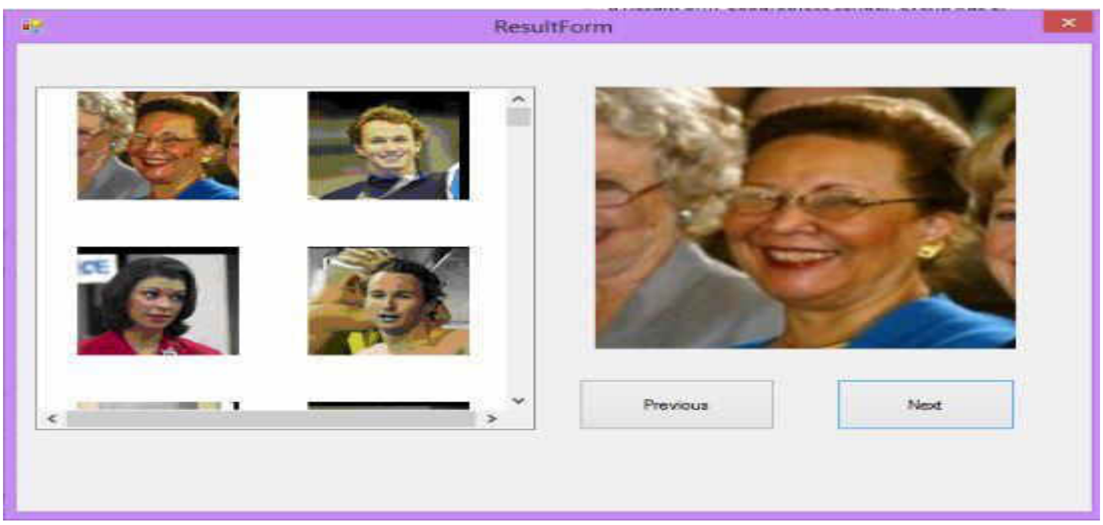

Figure 7: Emotion Detection System with its retrieval results

Performance Analysis of Facial Expression using Sobel Filters. Below Table shows different facial expression and its accuracy in this system where as for each expression the number of images is 60 and that accuracy is in percentage and total number of images in all over system is 240.

Facial Expression	Correct / No. of Images	Accuracy
Smile	46 / 60	76.66 %
Sad	32 / 60	53.33%
Surprise	40 / 60	66.66%
Normal	42 / 60	70%
Total	160 / 240	66.66%

5. CONCLUSION:

We proposed reliable emotion detection approach based on sobel filtering and bezier curve. The advantages of our approach are that a tilted human face can still be detected robustly even if the face is shirred, under shadow, of a different scale, under bad lighting conditions, and is wearing glasses. Thus this method can achieve a high-performance level in detecting human faces and extracting facial features in complex and simple backgrounds. This System can achieves accurate performance with 66.66%.

REFERENCES

1. Vladimir Vezhnevets, Vassili Sazonov and Alla Andreeva, "A Survey on Pixel-based Skin Colour Detection Techniques", Proceeding of the GraphiCon, pp.85-92, 2003.
2. Habibizad navin1+, Mir Kamal Mirnia : "A New Algorithm to Classify Face Emotions through Eye and Lip Features by Using Particle Swarm Optimization".2012.
3. Yong-Hwan Lee, Wuri Han, Youngseop Kim and Cheong-Ghil Kim "Robust Emotion Recognition Algorithm for Ambiguous Facial Expression using Optimized AAM and k-NN".
4. Rohini Patil, C.G.Patil, :"Image Processing Based Automatic Emotion Recognition And Classification Using Genetic Algorithm", E & TC Department, SAE, Kndhwa, Pune.
5. Yong-Hwan Lee, Yukong Lee : "Emotional Recognition System using Curve Fitting with Bezier Cubics.
6. Rein-Lien Hsu, Mohamed Abdel-Mattaleb and Anil K. Jain, "Face Detection in Colour Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. vol.24, no.5, pp.696-701.
7. Nicu Sebe, Ira Cohen, Theo Gevers and Thomas S. Huang, "Multimodal Approaches for Emotion Recognition: A Survey", Proceedings of SPIE – International Society for Optical Engineering, vol.5670, 2005.
8. Ira Cohen, Ashutosh Garg and Thomas S. Huang, "Emotion Recognition from Facial Expressions using Multilevel HMM", Neural Information Processing Systems, 2000.
9. Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable face image retrieval with identity-based quantization and multi-reference re-ranking," IEEE Conference on Computer Vision and Pattern Recognition, 2010.
10. N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Describable visual attributes for face verification and image search," in IEEE Trans- actions on Pattern Analysis and Machine Intelligence (PAMI), Special Issue on Real-World Face Recognition, Oct 2011.
11. N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," International Conference on Computer Vision, 2009.
12. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Conference on Computer Vision and Pattern Recognition, 2001.
13. Yoon W.-J. , Park K.-S., A Study of Emotion Recognition and Its Applications, 4th International Conference, MDAI, Springer, 2007.
14. Yoon W.-J. , Park K.-S., A study of speech emotion recognition and its application to mobile services, 4th International Conference, UIC, Springer, 2007.
15. Human Emotion Detection from Image by shakil0304003, 20 Sep 2010 CPOL.