Method for Face-Emotion Retrieval Using A Cartoon Emotional Expression Approach*

Vlaho KOSTOV**, Hideyoshi YANAGISAWA**, Martin JOHANSSON** and Shuichi FUKUDA**

A simple method for extracting emotion from a human face, as a form of non-verbal communication, was developed to cope with and optimize mobile communication in a globalized and diversified society. A cartoon face based model was developed and used to evaluate emotional content of real faces. After a pilot survey, basic rules were defined and student subjects were asked to express emotion using the cartoon face. Their face samples were then analyzed using principal component analysis and the Mahalanobis distance method. Feature parameters considered as having relations with emotions were extracted and new cartoon faces (based on these parameters) were generated. The subjects evaluated emotion of these cartoon faces again and we confirmed these parameters were suitable. To confirm how these parameters could be applied to real faces, we asked subjects to express the same emotions which were then captured electronically. Simple image processing techniques were also developed to extract these features from real faces and we then compared them with the cartoon face parameters. It is demonstrated via the cartoon face that we are able to express the emotions from very small amounts of information. As a result, real and cartoon faces correspond to each other. It is also shown that emotion could be extracted from still and dynamic real face images using these cartoon-based features.

Key Words: Design Engineering, Human Engineering, Human Interface, Face Impression, Cartoon Face, Emotion Extraction, Mahalanobis Distance, Principal Component Analysis

1. Introduction

The emotional state of humans is a subject of research in the domains of Kansel Engineering, psychology, medicine and is also a sociological and philosophical issue. However as globalization increases, more and more engineers are exposed to a world out of their own "societies", and forced to work and collaborate with different nation engineers across the borders. Getting objective measures of the way our colleagues think, understand, feel about something, or what impression they receive for the discussed issues will certainly help in minimizing misunderstandings and smoothing the collaboration of production processes in today's mobile and dynamic society. Being able to simplify the way the human "understands" other emotional states and reacts (as an impression reaction), enables us to model and improve any kind of interaction or communication process.

This paper presents an algorithm for a cartoon and real face impression analysis (face sketch and photograph of a real face) based on physical parameters. The significant finding on which it is based is that there is a relation between the cartoon faces and the real faces. We show that the cartoon faces are more appropriate representations regarding emotion (in terms of simplicity and speed). First, a cartoon face based emotional space was developed and determined, and real faces are mapped into it via a bi-directional link (Fig. 1). The same real face and cartoon face parameters are used to statistically determine the emotional group of the subjects.

We can use our system to send/receive emotion of a real face over distance and space (as a non-verbal communication). By our research, we have shown that we can apply the sketch face results on the real
face. We use it as a kind of emotional coding of the real face so we can send 9 parameters (9 simple numbers) as an emotional message over the network rather than significantly more parameters. This is important especially in the case of mobile equipment (telephones, hand held devices, organizers etc.) as the CPU power is small, and the network band is limited. Therefore we developed a method for dealing with emotion in the real face by dealing with simple numbers. Using such information, we are able to display the emotion and emotional atmosphere quickly and simply anytime and anywhere, without the need to send a whole face picture. It is a kind of emotional transfer or emotional compression if you like.

On the other hand, while technological advances may improve the image processing computational power problem, man machine interfaces using real face animations are facing some other simple obstacles. As the realism of the face increases (making the real face interfaces or agents look more like the real person), we become much less tolerant to imperfections in the modeling. Recent studies on responses of synthesized talking faces displayed on a computer screen reported that incautiously adding human characteristics like face, voice, and facial expression can make the experience worse rather than better\(^{[4]}\). If something looks like a real person we also expect it to behave as a real person because humans in every day life are able to read and perceive very subtle changes in real facial expressions\(^{[5]}\). Creating virtual creatures or characters that have non-human characteristics is a good approach (cats, dogs, virtual characters, cartoons)\(^{[2],[3]}\). In such a way we become less sensitive to the imperfections in the animation because we have no experience in talking or interacting with them. Therefore simple lines, or circles and curves can express or convey as much information as a complex 3D facial model. That is exactly what our cartoon face model is about.

2. Cartoon Face Analysis, Basic Method

At the very beginning of our research we had to consider a few important questions. What kind of faces are the most appropriate for emotion impression study, which emotional states should be considered, which face parameters are most emotion sensitive and relevant etc.? In order to solve these problems and to make the model as simple as possible we propose a face generation and evaluation model in Fig. 2. The emotion cartoon face space we obtain is optimized to fit the way the target group (students) understands, reacts and imagines emotions in faces. First we conduct a pilot survey as a means of face prototyping. We simply scan their ability or way of perceiving impression from emotional faces. By asking them to draw, we learn what is emotionally relevant for them, and how they imagine the emotional content of the cartoon face. After the second soft-rule based cartoon drawing, by feeding back the filtered design to the same students (we present them similar objects to be evaluated), we create emotion space that consists of subspaces and is optimized to their character. In our research, the target group was undergraduate.
students between the ages of 20 - 21.

2.1 Cartoon face space generation

53 students (45 male, 8 female) were asked to draw 9 different emotional states of a simple face (angry, happy, normal, sad, tired, scared, surprised, in trouble and stressed face) Fig. 3. The basic drawing rules were defined and the circle shaped face border was already printed. They were allowed to use straight lines, with different lengths and angles, and ellipses or circles with different heights, widths and angles. By doing this we actually extracted the way our target group understands emotion in face expression. It is especially powerful in emotional feature extraction since most humans easily express themselves by drawing, rather than by verbal interaction. The characteristic types of the faces, as patterns or sub-groups of similar faces that appear inside same emotion group were defined. The originally collected 477 cartoon faces from 9 different emotions, were given to 20 subjects, of different age and sex, for similarity evaluation. Inside each emotional group there were 53 different face drawings and the subjects were asked to group each face from a same emotion group into 6 different sub-groups according to their similarity. 6 sub-groups were chosen based on the famous Miller's work where he points out that on the average, our immediate memory can hold at most seven items. The faces that were repetitive between different sub groups were eliminated and by taking the mean representative from each sub group (the face that is closest to the group’s center of gravity) we formed the cartoon space of only 54 faces (9 emotion × 6 types × 1 face). With such approach, we formed 6 different types of same emotion drawings. Figure 4 shows 6 different characteristic types of

Angry and Happy faces. In case of happy face, the groups differ in the shape of the eyes (round eyes, ellipse eyes or straight-line eyes) as shown on Fig. 4 (b).

This “small” finite cartoon face space (54 faces) is a good representation of the 9 emotions we are analyzing. In the later sections we will show that it is possible to represent each of our 9 emotions by those cartoon faces and to obtain correct real-face emotion classification results.

2.2 Cartoon face parameters and emotional group generation

Nine physical parameters from each of the 54 cartoon face space are extracted, measured and computer generated and is shown in Fig. 5.
- Eye length (width) and height (l_e, h_e)
- Distance between eyes (d_e)
- Eye angle (A), Mouth angle (β)
- Eye distance to middle point of the face (y_e)
- Mouth distance to middle point of the face (y_m)
- Mouth length (width) & height (l_m, h_m)

From these parameters we created a 54 × 9 PM matrix (Parameter Matrix) and is shown in Fig. 6. Those 54 faces are then evaluated by the target group (students) for emotion impression evaluation. Since we require that the subject must be aware of all gradation at the time and we agree with the psychological experiments, which show that an individual can not simultaneously compare more than seven objects (plus or minus two) without being confused, we chose 7+2=9 emotions for evaluation. Each of the 54 faces is evaluated by each student on a scale from 1 to 5 for the same nine emotions that they were asked to draw. In such a way we establish and keep the consistency of the target group evaluation. As a result we obtain the Emotion matrix (EMAT) shown in Fig. 7, which contains the emotional scoring for each face per student. In order to gain the evaluation accuracy of our model, an Impression Web based survey was conducted. However, an enormous expansion of examinee numbers did not change the results significantly. It was proven rather, that the number of emotional space members mostly affects the accuracy.

![Fig. 3 Samples of cartoon faces](image)

![Fig. 4 Angry and Happy 6 characteristic types of faces](image)
2.3 The algorithm

As a first step we do preprocessing of PM by a standardized Z score. It returns the deviation of each column of PM from its mean, normalized by its standard deviation. In order to form new variables \( P_{\text{new}} \), which are linear composites of the original variables, we perform Principal Components Analysis (PCA) given by Eq. (1):

\[
P_{\text{new}1} = a_{11}l_e + a_{12}h_e + \cdots + a_{19}h_m + a_{10}\beta \\
P_{\text{new}2} = a_{21}l_e + a_{22}h_e + \cdots + a_{29}h_m + a_{20}\beta \\
\vdots \\
P_{\text{new}9} = a_{91}l_e + a_{92}h_e + \cdots + a_{99}h_m + a_{90}\beta
\]

where \( a_{ij} \) are the principal components weights, and \( l_e, h_e, h_m, \beta \), etc. are the original 9 parameter data (eye_length (width), eye_height, mouth_height, mouth_angle etc.). In our research we take the PCA components that account for 99% of the total variance of the original data. In such a way we obtain dimension reduction (in our case we reduce the number of variables). PCA is used for two main reasons: to break the data in its most basic variations (eigenvectors or principal components) and to be able to perform Mahalanobis Distance that in some cases performs a limitation of 10–15 variables. Another limitation is that the number of samples in the teaching data should be at least as big as the number of variables we use. In the case of all members of the teaching data having a variable that doesn't change, the covariance \( S \) is zero and we are unable to perform the calculation Eq. (2). For the \( p \) variable case, the Mahalanobis distance between observations is given by Eq. (2).

\[
MD_{ij} = (x_i - x_k)S^{-1}(x_i - x_k)
\]

\( x \) is a \( p \times 1 \) vector of coordinates and \( S \) is a \( P \times P \) covariance matrix. We used the first 6 eigenvectors \( (P_{\text{new}}) \) that satisfied the 99% condition. EMAT was averaged by its 3rd dimension, taking the mean value of each student evaluation. A 54 × 9 matrix was obtained Eq. (3).

\[
EMAT_{ij} = \sum_{k=1}^{n_s} EMAT_{ik}/n_s
\]

where \( n_s \) is the number of subjects (for this case \( n_s = 53 \)). We calculate the emotion reference space for each emotion according to the values of EMAT. We extract the first 9 faces per emotion with the highest score in the impression evaluation Fig. 8. Some of the face overlap in fact belongs to different groups at the same time. However that creates smooth borders between each emotion reference space. Even in the real world, the same faces may look partly stressed while in trouble, and some may look partly sad while normal. Up to 9 faces per emotion could be considered as relatively good representatives for that particular emotion, but the last 3 faces appeared very often in the other emotional spaces as well. Therefore after performing PCA and considering the first 6 components, the Teaching Data (TD) (Fig. 9) was obtained by taking only the first 6 faces per emotion. Then the face data processing is done. We bring a new sample \( (\text{data}_{\text{newdata}}, 1 \times 9 \text{ vector}) \) in our cartoon emotional model, first normalizing it with PM’s mean and variance. Then PCA is performed and finally the Mahalanobis distance between the new sample and each Teaching Data (TD) set is calculated by Eq. (4).

\[
MD_{i} = (\text{data}_{\text{newdata}} - \bar{a}_i)S^{-1}(\text{data}_{\text{newdata}} - \bar{a}_i)^T
\]

where \( \text{data}_{\text{newdata}} \) is new face vector \( [1 \times 9] \), \( \bar{a}_i \) is the average of the teacher data for emotion \( i \) \( [1 \times 9] \), \( S \) is the covariance for emotion \( i \). The new sample belongs to the emotional group from which it has
minimum distance \( MD_i = \min \). Since our training set consists of a small number of data, each reference space consists of only 6 or 9 faces that are quite different from the other reference spaces. Therefore, while calculating the Mahalanobis distance Eq. (4) we use the covariance matrix \( S_i \) of the whole PM training set rather than the covariance of that particular reference space.

3. Real Face Analysis

A low-resolution \((240 \times 320)\) Web videoconference camera was used to capture images of 12 individuals (8 Japanese, 4 non-Japanese). Each person was asked to perform a face gesticulation according to the previously mentioned nine emotional states. We have constructed a real face space that consists of 6 images of each emotion (54 real face images). The same subjects (53 students) evaluated the real images through a written survey in the same way they did with the cartoon faces. As a result we constructed the Real face Emotional Matrix (REMAT) that was generally used for model verification. The physical parameter matrix (RPM) is extracted by a custom image processing algorithm. One of the design goals of the algorithm was to keep it as simple as possible, while still making it useful. We use a uniform background (black or dark) because our aim was not to develop an ultimate image processing face extraction algorithm. We use it in order to extract the physical parameters (mouth length/height, eyes height/length etc.) For that purpose our algorithm works reasonably well enabling us to show the uniqueness of our idea. A literature survey of the existing research in this area shows that if we compare the complexity and results obtained, against our simplicity and results obtained we believe our approach is better. The algorithm can roughly be divided into the 4 following steps:

- Segmentation by thresholding
- Extraction and ellipse fitting of the face.
- Extraction of the facial features
- Measurements

3.1 Segmentation

The first step is to try to separate the face from the background. Since this is currently an active
research area\(^{40-49}\), we decided to simplify the approach by only using a dark background. With this limitation, it is possible to separate the face from the background using a simple adaptive thresholding scheme. But, since the pictures were taken under non-uniform lighting conditions, we have to try and compensate for this. Since our only concern at this stage was to separate the face from the background, we simply subtract the background from the picture. We get an approximation of the background by interpolating from the edges of the picture, using simple linear interpolation (first from left to right, and then up to down). This gives us a new picture with an almost uniform background. The next step is to calculate the value we should use as a threshold. This is accomplished by looking at the probability distribution function (pdf) for the pixel brightness. A typical pdf is shown in Fig. 10. The background corresponds to the first huge peak, since it contains a lot of dark pixels. The remainder is the face. So we then calculate the place where this peak levels off, that is when the derivative is small. Since we are using a discrete derivative, it will not be zeros when there is an extreme value, and since the peak is very sharp and narrow, our approach works.

3.2 Extraction of the face

In order to separate the face from the hair and shoulder etc., it is necessary to find the edges of the face. This is easily done by morphological operations. First, erosion is performed on the binary picture, from the previous stage Fig. 11(a). When the eroded picture is subtracted from the binary image, the result is an image showing only the edges Fig. 11(b). This kind of picture is hereafter referred to as an edge map. The edge showing of the face is then simply the longest edge in this new picture Fig. 11(c). But this edge is always connected with other edges, the shoulder, the hair, etc. So these artifacts need to be removed, if the next stage in the algorithm should work. This is done by placing the origin of the coordinate system at the center of the face, and converting to polar coordinates Fig. 11(d). Since the face forms an ellipse around the origin, it can be extracted by taking the pixels closest to the origin in every direction. This is easily done in polar coordinates. The result is converted back to Cartesian coordinates, and the edge of the face is all that remains, Fig. 11(e). The last part of this step is to fit an ellipse to the face. This is done by a least-squares method developed by Fitzgibbon et al.\(^{91}\) The ellipse gives us a convenient way of obtaining the orientation of the face, and to divide the face into the upper and lower regions, Fig. 11(f). Various techniques using edge-detectors and gradient methods to segment the image and extract the face prior to fitting the ellipse were tried, but the latter one worked better. It also has the advantage of being very simple and easy to understand. We also tried to locate the nose and to use that for correct rotation, but it did not improve the performance.

3.3 Extraction of eyes and eyebrows

Since there is no constraint placed on the pictures concerning the orientation of the face, some of the pictures contain faces that are tilted. The first thing we do is to rotate these pictures so that the face is straight. Then the picture is cropped keeping only the central part as defined by the user. We do this with both the edge-map, and the original image. In order to locate the eye and the eyebrow first a morphological dilation is performed on the edge-map. After this, a mirror image of the edge-map is constructed, and a logical “and” between the edge-map and its mirror is performed. Since the placement of the eyes and the eyebrows are symmetrical with respect to the axis of the face, this gives us a new picture with blobs where the eyes and the eyebrows are located Fig. 12 (left). (A

![Fig. 10 Probability distribution function of brightness](image)

![Fig. 11 Extraction of face](image)
blob is an accepted term in image processing which means a connected collection of pixels. It looks like a blob, hence its name.) The dilation is used to guarantee that the edges overlap when the image is mirrored. After this step, we can look only at the left half of the image, since it is symmetrical. Once the location of the eye is known, it is extracted by a simple thresholding scheme. The eyelashes and the pupil are much darker than the average pixel value in the area. So we simply subtract the average and threshold. This gives us the eye. The eyebrow is extracted using a combination of gradients and thresholding. The edge of the eyebrow is located by calculating the magnitude of the gradient. The maximum value is on the edge of the eyebrow Fig. 12 (right). This is then used to select the blob corresponding to the eyebrow. The blob is obtained by thresholding.

3.4 Extraction of the mouth

Due to the fitting of the ellipse, we know the location of the lower part of the face. So the first step is to crop the image, and only keep this part. The mouth itself is then located by looking at the first and second order derivative in the x-direction. A high positive value of the first order derivative corresponds to the upper edge of the upper lip, while a high positive value of the second order derivative, corresponds to a dark valley in the picture. This is the area between the lips. Therefore, we use the first order derivative to locate the mouth, and the second order derivative as an estimation of the mouth. Once again we must threshold the second order derivative in order to locate the mouth Fig. 13. We also tried other standard edge detection techniques (Sobel method, Prewitt method, Roberts, zero cross method and Laplacian of Gaussian Method), and simple thresholding. But we found out that our method works better. The other approaches often mistook the nose for the mouth. Our method therefore is the most robust approach used so far.

3.5 Face features measurement

Once the extraction is completed, it is quite trivial to measure the necessary parameters. From the real faces, we extract exactly the same 9 parameters we extracted from the cartoon faces, Fig. 12 (right), Fig. 13 (right), plus the eyebrow parameters. The eye-angle is obtained by locating the left- and rightmost pixels of the eye. The slope of the line between these pixels gives the angle. After rotating the eye, so the angle is zero, the width and height are easily determined. The eyebrow is measured in the exact same way as the eye. Implementing logical conditions for the obtained eye and eyebrow parameters, we correctly map them into the cartoon face space parameters.

The mouth is more complex, since it is much more mobile. The strategy we chose was to locate four points: One point on the middle of the upper lip, one on the middle of the lower lip, and the two points at the corner of the mouth. These points are then used by some simple empirical rules to determine if the mouth is open, if it is a straight line etc. After that, it is trivial to calculate the desired parameters from the four points.

4. Emotion Expression Analysis

There are many different approaches in choosing the appropriate emotion categorization. Most research in facial expression recognition is based on the Ekman’s six basic emotion categorizations that are claimed to be recognized across all cultures\(^{90}\) (joy, fear, anger, disgust, sadness and surprise). Kansel Engineering researchers defined emotions into pure and mixed states\(^{111}\). Some of them are based on Plutchik theory that states there are 4 basic pairs of emotions called pure emotions and all other emotions are derived from those\(^{111}(13)\) (Sadness-Fear, Hatred-Surprised, Expectation-Acceptance, Anger-Joy). Others use IAPS (The International Affective Picture System)\(^{112}\), to induce seven target emotion conditions (anger, disgust, fear, happy/calm, happy/excited, neutral and sadness). Others\(^{113}(14)\) use the FACS, The Facial Action Coding System by Ekman and Friesen\(^{17}\) composed of 44 “action units” that in combination represent all visible discriminable face expressions.

However, in our case we try to represent the emotional state of the human real face in a somehow different manner. We applied a modified Ekman’s
categorization, expanded by few more categories: normal, tired, in trouble and stressed. We concentrate on the final emotion state impression detection and how we perceive the emotion rather than the emotion theory and derivations.

4.1 Emotion reduction

During the process of taking the camera snap shots, most of the participants suggested that some of the emotional states they were asked to perform were either too similar, or too difficult to discriminate and/or imitate. Therefore we tried to group the emotions according to the way the subjects perceived them. Figures 14 and 15 show the dendrogram plots (hierarchical binary cluster tree) of the cartoon face and real face emotional matrices (EMAT and REMAT).

They are computed as Single Linkage, Euclidean distances between pairs of objects in the data. The Cophenetic correlation coefficient, that measures the distortion of this classification indicating how readily the data fits into the structure suggested by the classification, is 0.8120 and 0.5402 respectively. By observing the real and cartoon face plots of amalgamation schedules (graph of the linkage distances across consecutive steps of the linking process) on Figs. 16 and 17, we can identify plateaus where many clusters are formed at approximately the same linkage distance. We can also indicate a natural “discontinuity” in terms of distances between the observed objects so we conclude that cartoon faces sad/tired and scared/in trouble may be considered as one emotional state. In the case of real faces, it applies for tired/stress and sad/in trouble. By observing the natural alignment of the data, we can find the natural divisions in the data set and conclude that in both cases (cartoon and real face) we could divide the cluster tree into only 5 emotional states. 1: angry 2: happy 3: normal 4: sad/tired/scared/in trouble/in stress 5: surprised.

The face emotion classification accuracy is: in the case of 9 emotions 70% correct; 23% close and 7% error. In the case of 7 emotions (1: angry 2: happy 3: normal 4: sad/tired 5: scared/in trouble 6: surprised 7: in stress), 76% correct, 18% close, 6% error. In the case of 5 emotions, the emotion recognition rate is 83% correct 13% close, 4% error. If we reduce the emotions from 9 to 5 we gain on speed and we reduce the computational time by 44.4%. In the case of 9 to 7 emotion reductions the computational

![Fig. 14 Dendrogram for cartoon faces, single linkage, euclidean distance](image1)

![Fig. 16 Plot of linkage distances across steps (euclidean distances for cartoon faces)](image2)

![Fig. 15 Dendrogram for real faces, single linkage, euclidean distance](image3)

![Fig. 17 Plot of linkage distances across steps (euclidean distances for real faces)](image4)
time is decreased by 22.2%.

It is shown that 9 emotions could be reduced by looking at the respective binary trees (dendrograms) in order to increase the emotion recognition accuracy, and save computational time. In the case of real faces we could (not necessarily should) reduce it to 7 or 5 emotions. By using a system with 9 emotions we also have shown that there are rules applicable in general to all faces. Example: real faces "sad" and "in trouble" could be considered as one emotional state because they are evaluated in same manner (shown by REMAT, emotion matrix). If we would like to go in more detail and discriminate between those two, we should look in the details that separate those two emotions. This could be easily accomplished by looking at the Correlation maps (emotion versus physical parameters).

4.2 Emotion and physical parameter correlation

The correlation matrix plot between emotion and physical parameters of the cartoon face (EMAT and PM) is given on Fig. 18. The eye center distance (center of eye to center of face) and eye distance has the smallest correlation with any of the emotions. Therefore we take the real face eye-eye distance as a referent one for re-scaling of all other face parameters. In such a way, we solve the real face scaling problem (when some of the real face pictures are zoomed our approach still works). The eye-angle parameter has the strongest influence on almost all emotion groups. In the case of the real face, the correlation matrix between the emotion and physical parameters (REMAT and RPM) is given by Fig. 19. In this case, the eye-angle influence is minimized and the eyebrow angle took its role. This is because when people draw faces, they normally create an exaggerated version of a real face. Since the cartoon model does not contain eyebrows, the students could draw faces with only lines or ellipses instead of the real eyes. In some of the emotions they refer to the eyebrow (angry, sad) but in some cases they refer to the eyes (normal, happy, etc.). In fact if we replace the eye-angle of the cartoon face with the eyebrow of the real face, we have the same high correlation of the physical parameters for the state of scared, angry, normal and surprised. Disregarding eyebrows in the cartoon model is simple and powerful method contributing towards our goal of using simple cartoon face models for the most relevant physical parameter extraction of emotions. Therefore we conclude that the eye and eyebrow issue is a specific one and even though we could consider them in a somewhat different manner, they have much in common.

The subjects express themselves (by drawing), keeping in mind that they had to draw a face(s) as simple as possible. Of great importance is the fact that we didn’t actually decide the rules of the model a priori, but rather we extracted them from the subjects by emotion face prototyping. We asked 20 subjects in a pilot survey to draw "free" emotional faces and the drawing rules for the actual survey were defined following that cartoon face prototyping.

For cartoon faces, at the part of the eyes only lines or ellipses were used. We prove that this simple line (in the cartoon model referred to as eye parameter), in some cases, means eyes, but in other cases means eyebrow. We can call it “eye region” parameter rather than “eye parameter”, because it represents both the eye and eyebrow at the same time, depending on which is more important for that particular emotion. We actually use only one parameter to represent two actual features or objects of the real face (which seems to be good parameter reduction). Example: In case of the normal face, the cartoon model eye angle parameter means the real eye angle, because it’s a horizontal line (eyebrow doesn’t change). In the case of angry, the inclined line represents the eye-brow.

Fig. 18 Correlation map between emotion evaluations and feature parameters for cartoon faces

Fig. 19 Correlation map between emotion evaluations and feature parameters for real faces
(the eye doesn’t change). In the case of surprised, the eye-height of the cartoon model is the same as the real eye-height.

Knowing the eyebrow importance, at the real face analysis we extracted the eyebrow and have them compared with the eyes from the cartoon face. We showed that the eyebrow angle of the real face has the same significance as the eye-angle of the cartoon model. However, the eyebrow width and length change very little per emotion at the real face.

5. Real Face Emotion Retrieval

Even though many applications require detailed emotion extraction, some applications require speed i.e. fast and basic emotion detection. Our method is powerful enough to deal with both cases in any particular application and with its unique approach it achieves the accuracy of the previously developed methods while being computationally inexpensive. The final idea of this work is to develop a system that could detect the users’ emotion in reaction to interaction with a machine or with another person over a network.

5.1 Static real face evaluation (Impression evaluation of real face emotion)

As our aim was to show that the impression groups developed from the cartoon face analysis can do real face impression evaluation, we tested the system with 30 still, real face images from different individuals (23 were used for the real face survey, and 7 were completely new). First the real face image was converted to its a simplified cartoon representation and later evaluated by the cartoon emotion grouping system as shown in Fig.20. Our model correctly evaluated 72% of the cases and 17% were associated

![Figure 20](image1)

Fig. 20 Procedure for emotion evaluation by feature parameters

![Figure 21](image2)

Fig. 21 Dynamic emotional face analysis
to a similar emotion.

5.2 Dynamic real face evaluation

In order to extract emotion dynamically, we applied the same algorithm on the still images (video frames) captured from a face motion video at each $\Delta t$ time intervals. Since we wanted to extract emotions or affect from the human face, which do not change so rapidly, based on our experiments we proposed a $\Delta t > 1$s. Simple emotional content video analysis of a Web videoconference session clearly shows that high frame rate would be nothing but redundant data. One segment of our dynamic emotional face analysis is shown on Fig. 21. Each still image captured from the video is shown with its corresponding cartoon face and the emotion evaluation. Physical face parameter change is dynamically associated to the physiological change of the emotion and/or mood (Emotional difference, (Emotion), which opens the field to various applications requiring fast and simple face emotion extraction.

Our target in this work was a post processing of video data, rather than real time application. In case of real-time application it would be necessary to develop embedded systems along with more efficient algorithm for detection of emotion from video. The change of the affections (or mood/emotions) directly from the changes of the parameters between the neighboring snapshots can be used to speed the analysis. However, it was out the scope of this paper as the real time application and the speed was not a criterion for the Video extraction part in our research. The physical parameters that were derived from the cartoon face model were proved to be minimum number of parameters that can successfully represent emotion in human face. Therefore, the future extension of this work could include a research for efficient coding of affection in real face based on the cartoon face parameters.

6. Conclusion and Discussions

Modern society will become more complex and increasingly global ultimately requiring communication methods that will include the ability to determine human emotion. The human face is one of the most important objects for emotion extraction.

In this paper a unique and simple approach towards understanding of emotion is introduced. It is shown that it is possible to extract emotion from a real face by using a cartoon emotional expression approach. The emotional space developed for the cartoon faces was used to evaluate real faces by simplifying them to their basic features and mapping them to the cartoon face space.

Our work is towards sending and detecting emotion over network, including detection of the face emotion of the party at the far end. In a case of reduced channel capacity in the communication via the low affective bandwidth of e-mail, users have indeed invented smiley, so-called emoticons such as (",;:) and (; to be able to convey minimal emotional coloring in a message. However, those are used to send the intended emotional message only, i.e. what we would like to make the others believe. On contrary, with our approach we can extract the real face emotion in addition to the emotional message thus we are able to bring the face-to-face experience into computer-mediated communication without even sharing the same physical space. We started from Cartoon Model and we were successful in reducing the parameters. We strongly believe we are on the right track with our approach of extracting emotion from real face based on a previously developed cartoon model, because it successfully represents the emotion (or affect) with very small number of parameters.

We might argue that in some cases the cartoon faces might be exaggerated representation of the real faces and that there are cultural and sex dependencies. However, those arguments are reflected into the final classification results. More work could have been done in developing and tuning the statistical methods to measure the “exaggeration rate” and accordingly adopt and improve the classification results, but that would have been outside the scope of this paper. Our primary goal was to show this novel approach of representing emotion by cartoon and then analyzing a real face.

Face-to-face meetings are still the most popular forms of business communication. 44 percent of executives surveyed by office staffing service, said they preferred to meet with people in person, E-mail ranked second at 34 percent; paper memos, 12 percent; and voice mail, 7 percent. In these days of rapid change, time is precious, and we can’t afford to waste it in conventional meetings, especially when our partners or colleagues are on the other side of the globe. Asynchronous communication will be more and more common and having a simple and fast affective interface enabling us face-to-face experience is essential for better communication.

The cartoon face impression expression was categorized using PCA and Mahalanobis distance methods and face feature parameters were determined and related with each emotion. Regarding the face image analysis, we developed an image processing method that is able to extract the same parameters easily. Applying the emotion evaluation method that was developed and based on cartoon faces we confirmed that it is applicable to the real face. We
have shown that the impression we perceive by looking at the face sketches is the same that we would get from a real face. Our research was divided into a few steps. As a first step the cartoon face research was completed and as a second step the real face analysis (static image processing) was done. The third step was emotion evaluation and feature extraction implemented on static and dynamic image using a slow frame rate (1 frame/sec). Changes of physical face parameters are detected dynamically and associated to the physiological change of the emotion and/or mood.

Up until this point of our research we successfully showed the concept through simplified approaches. The change of the affections (or mood/emotions) directly from the changes of the parameters between the neighboring snapshots can be used to speed the analysis. Implementation of Fuzzy c-means clustering method, or Adaptive Resonance Theory and the development of the optimized methods for an effective real time systems would be the target of our future extension of this research.

References