International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

# A New Era of Emotion Recognition

Mahjabeen Khan[A*], Sumitra Thombre[A], Priti Ingle[A], V.R.Gosavi[A] and A. K. Deshmane[B]

[A]G.S. Mandal's, Maharashtra Institute of Technology, Aurangabad (Maharashtra), India
[B]JSPM'S, B.I.T, Bharshi, India

Accepted 15 Nov 2014, Available online 25 Dec 2014, **Vol.4, No.6 (Dec 2014)**

***Abstract***

*Emotion of a person plays important role in life because we cannot express our feelings or emotions in words, facial expression or gesture to express emotions. A human face does not only identify an individual but also communicates useful information about a person's emotional state. Facial Expression gives important information about emotion of a person. Face emotion recognition is one of the main applications of machine vision that widely attended in recent years. It can be used in areas of security, entertainment and human machine interface (HMI). Emotion recognition usually uses of science image processing, speech processing, gesture signal processing and physiological signal processing. In this paper a new algorithm based on a set of images to face emotion recognition has been proposed. This process involves four stages pre-processing, edge detection, feature extraction, face detection.*

***Keywords:*** *Color Space, Edge Detection, Face Detection, PCA*

## 1. Introduction

*Overview*

The area of human-computer interaction (HCI) will be much more effective if a computer is able to recognize the emotional state of human being. Emotional states have a greater effect on the face which can tell about mood of the person. So if we can recognize facial expressions, we will know something about the human's emotions and mood. The objective of this research is to develop Automatic Facial Expression Recognition System (AFERS) which can take human facial images containing some expression as input and recognize and classify it into appropriate expression class such as angry, disgust, fear, happy, neutral, sad, and surprise.

This research focuses on the investigation of computer vision techniques designed to increase both the recognition accuracy and computational efficiency by applying some modifications in terms of face localization, feature extraction and classification algorithms and hence arriving at a simpler approach to perform facial expression recognition and classification. Faces are accessible windows into the mechanisms which governs our emotional and social lives. About 70% of human communication is based on non-verbal communication such as facial expressions and body movements. In 1872, Darwin wrote a treatise that established the general principles of expression and the means of expressions in both humans and animals. He also grouped various kinds of expressions into similar categories.

*Corresponding author **Mahjabeen Khan, Sumitra Thombre** and **Priti Ingle** are students; **V.R.Gosavi** is working as Assistant Professor and **Dr. A. K. Deshmane** as Principal

*Facial expression*

Facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. According to Fasel and Luttin , facial expressions are temporally deformed facial features such as eye lids, eye brows, nose, lips and skin texture generated by contractions of facial muscles. They observed typical changes of muscular activities to be brief, lasting for a few seconds, but rarely more than five seconds or less than 250 ms..


**Figure No.1**

They also point out the important fact that felt emotions are only one source of facial expressions besides others like verbal and non-verbal communication or physiological activities. Though facial expressions obviously are not to equate with emotions, in the computer vision community, the term facial expression recognition often refers to the classification of facial features in one of the six basic emotions: happiness, sadness, fear, disgust, surprise and anger. Facial expressions play an important role in our relations. They can reveal the attention, personality, intention and psychological state of a person . They are interactive signals that can regulate our interactions with the environment and other persons in our vicinity. According to Mehrabian, about 7% of human communication information is communicated by linguistic language (verbal part), 38% by paralanguage (vocal part) and 55% by facial expression. Therefore facial expressions provide the most important information for emotions perception in face to face communication. The six basic facial expressions are as shown in Figure 1.

*Automatic Facial Expression Recognition (AFER) system*

Automatic facial expression recognition (AFER) System is gaining an interest in various application areas like lie detection, neurology, intelligent environments, clinical psychology, behavioral and cognitive sciences and multimodal human computer interface (HCI) . It uses the facial signals as one of the important modality and causes interaction between human and computer in more robust, flexible and natural way. In surveillance system and in intelligent environment, AFER is useful in following ways:

1. A real-time automatic surveillance system which detects human faces and facial expressions accurately can be installed at busy public places like malls, airport, railway station or bus station around the world so that it can avoid the possible terrorist attack. The system would detect and record the face and facial expression of each person/ passenger. If there were any faces that appeared to look angry or fearful for a period of time, the system might set off an internal alarm to warn the security personnel about the suspicious passengers.
2. In a real-time gaming application, a real-time facial expression recognition system can observe players' facial expressions. If a player shows surprise or excitement, the system would know that the particular part of a game is being highly enjoyed by the player. If a facial expression appeared to be neutral for a period of time, the system might notify the game to change some of its elements or difficulty levels. This kind of intelligence can enhance playability and interactivity of different types of games.
3. In educational games like a math learning game for elementary school students could tell if the math topic shown on a screen is too difficult based on the facial expression of a student who is playing the game.
4. In driver observation system, a sleepy face of a driver can be traced by the camera and may indicate whether he or she is getting tired while driving. Then the system might set off some warning signals to the driver or be able to help the driver to pull over safely. Such a system might prevent many accidents caused by driving under the influence or fatigue of a driver.
5. In educational institutions, real time facial expression recognition system is useful to detect or record the expressions of the students sitting in a class. A teacher can evaluate himself from the recorded expressions and can modify his methodology of teaching.

## 2. Ease of Use

Development of fully automatic facial expression recognition system (AFERS) is a challenging and complex topic in computer vision due to various factors like pose and illumination variations, different age, gender, ethnicity, facial hair, occlusion, head motions, lower intensity of expressions and other difficulties. Facial expressions are generated by contraction or relaxation of facial muscles or by other physiological processes such as coloring of the skin, tears in the eyes or sweat on the skin. Facial expression represents a particular pattern. In order to classify/recognize a pattern into appropriate class there is a need to extract information from the patterns and produce feature values. Feature information is obtained in two ways:
1) Appearance based features-uses color/texture information about the image pixels of the face to infer the facial expression
2) Geometry based features- analyze the geometric relationship between certain key points (fiducial points) on the face when making its decision.

*Issues in FERS*

Many modern FER systems use the geometric positions of certain key facial points as well as these points' relative positions to each other as the input feature vector. Some researchers used real valued and binary parameters and distance parameters to extract facial features for expression recognition. Limitation of the above systems is that all have used manual approach for pointing or extracting features which was time consuming.

Many modern FER systems were proposed to recognize few of facial expressions out of seven basic facial expressions. These systems were unable to recognize all seven basic facial expressions. It has also been observed that the problem of facial expression recognition has been carried out mostly on the basis of comparison of other expression images with neutral images. This approach increases the complexity in terms of comparison which slows down the speed of computation. Also it increases the memory requirement of the system.

Many researchers used PDM/ASM like model based approaches for feature extraction. But these approaches were suffering from the fact that manual labor is necessary to construct shape models .Many modern FER systems use the appearance based features extracted using the techniques like LBP,wavelets,PCA,ICA,EICA,FLDA and achieved recognition accuracy in a moderate range for limited number of images.

Thus manually pointing the positions of feature point for feature extraction, manual constructions of model in ASM/PDM techniques, recognition of few of the facial expressions instead of recognizing all seven facial expressions and recognition efficiency are the major issues to be considered as far as existing AFERS are concerned.

*Motivation*

We were motivated to increase the speed of computation, better utilization of memory and to achieve high efficiency in terms of classification and recognition of facial expressions by suggesting some of the modifications in terms of feature extraction, classification and recognition algorithms.

*Objective*

To achieve high degree of efficiency based on the motivation and stringent requirement of improving accuracy and covering all expression classes, the research topic is selected as Classification and Recognition of Facial Expressions for Human Faces. The goal of this research is to apply some modifications in terms of feature extraction techniques and algorithms for classification and recognition, by using existing image processing operations hence arriving at simpler approach to perform facial expression classification and recognition which will improve the classification and recognition accuracy. To meet the estimated goals, the objective of this research is to develop Automatic Facial Expression Recognition System (AFERS) which can take human facial images containing some expression as input and classify and recognize it into appropriate expression class.

AFERS will automatically carry out
1. Preprocessing of facial images.
2. Localization of face portion required for feature extraction.
3. Extraction of facial features.
4. Classification and recognition of facial expressions using appropriate classifier

## 3. Face Recognition

*A. Mechanisms for Recognizing Emotion from Faces*

We begin with an outline of the different possible mechanisms for recognizing emotions from facial expressions. In the following sections, these possible mechanisms will then be tied to specific neural structures and their interconnections. One conclusion will be that a given brain structure typically participates in multiple strategies and that performance on a recognition task also often engages disparate strategies and, hence, disparate sets of neural structures.

*1. Recognition as Part of Perception*

One possibility is to consider recognition as a part of perception. Arguably, recognition of simple features of a stimulus, or recognition that one stimulus differs from another, is really an aspect of perception. Perhaps we do not need to know anything about the world to recognize an emotion but are able to discriminate, categorize, and identify emotions solely on the basis of the geometric visual properties of a stimulus image. It is even conceivable (in principle) that such perceptual processing could be linked directly to language-related regions of the brain sufficient to produce the name of the emotion, in the absence of retrieving any other information associated with the stimulus (something akin to paired associate learning, for instance).

*2. Recognition via the Generation of Associated Knowledge*

However, recognition typically involves more than just perceptual information. When we see a facial expression of fear, we can relate it not only to the percepts of other facial expressions in terms of its structure, but we can recognize that the person whose face we see is likely to scream, is likely to run away, has probably encountered something scary, and so on. None of that knowledge is present in the structure of the stimulus; it is present in our past experience with the world (and, to some limited extent, may even be present innately). A complex question concerns the precise mechanisms by which such knowledge might be retrieved. In general, the knowledge is not stored in any explicit format but rather relies on recipes for reconstructing knowledge by reactivation of the representations that were originally associated with one another when the knowledge was acquired (e.g., A. R. Damasio & Damasio, 1994). The simplest example of such a mechanism would be literal association, as when we see a face of fear and hear a scream at the same time and link the two henceforth in memory. In general, linking other knowledge with a perception of the facial expression will be vastly more complex and will rely on multiple higher order associations that may be fairly separated in time (e.g., seeing a face of fear and seeing the chasing tiger some time later), as well as on symbolic representations that, in humans, rely substantially on language (e.g., seeing a face of fear and merely being told that the person was afraid because he or she was running away from a tiger).

The general neural scheme for implementing the above mechanisms requires the binding of information between separate neural representations so that they can be processed as components of knowledge about the same concept. In the perceptual case, a stimulus activates multiple neural regions that represent particular aspects of its visual properties, and the coherent ensemble of these different bits of knowledge (the representations of the different properties of the stimulus) constitutes the perceptual mechanism that we discussed in Section 4 above. But, this mechanism can be extended beyond those neural regions that represent the visual properties of the stimulus to include those that represent knowledge not of the stimulus itself but of that with which it has been associated. The demand for integrating neural representations that are spatially separated in the brain would require extensive feedback connections as well as feedforward connections between different neural regions. One might thus envision a continuous dynamic interplay between feedforward, feedback, and horizontal information flow from which the brain constructs representations of visual stimuli (cf. Lamme, Super, & Spekreijse, 1998, for review). Schemes such as Ullman’s (1995) counter streams or Edelman’s (1987) re-entry both capture this idea: The representation of the stimulus itself, and of its associated knowledge, evolves contemporaneously such that the one continuously modulates the other and perception and recognition become parts of the same large-scale process.

*3. Recognition via the Generation of a Simulation*

The above mechanisms, although they rightly can be considered creative, are relatively direct: On linking together the various representations that give rise to components of the conceptual knowledge about the emotion that is signaled by the stimulus, the subject has available all the information necessary to recognize the emotion; all that is required to perform most recognition tasks now is an implementation of the reconstructed conceptual knowledge in terms of language so that the subject can tell us what he or she knows. But there are less direct routes that might come into play also. It may be that the explicit knowledge triggered in the above scheme is insufficient to recognize an emotion, perhaps because that particular emotion was never seen before or because the recipe for reconstructing knowledge about it provides insufficient detail. Another mechanism might attempt to generate conceptual knowledge using an inverse mapping that seeks to trigger those states normally antecedent to producing the facial expression. Such a mechanism would attempt to simulate in the observer the state of the person shown in the timulus by estimating the motor representations that gave rise to the observed stimulus. Once the observer has generated the state that the other person is presumed to share, a representation of this actual state in the observer could in turn trigger conceptual knowledge. Simulation thus still requires the triggering of conceptual knowledge, but the basis of the trigger is not a representation of someone else but rather a representation of ourselves (simulating the other person). The simulation hypothesis has recently received considerable attention due to experimental findings that appear to support it. In the premotor cortex of monkeys, Rizzolatti and colleagues have reported neurons that respond not only when the monkey prepares to perform an action itself but also when the monkey observes the same visually presented action performed by someone else (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Gallese & Goldman, 1999; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Various supportive findings have also been obtained in humans: Observing another's actions results in desynchronization in motor cortex as measured with Adolphs It thus appears that primates construct motor representations suited to performing the same action that they visually perceive someone else perform, in line with the simulation theory.

*The Development of Emotion Recognition*

We begin with an outline of the different possible mechanisms for recognizing emotions from facial expressions. In the following sections, these possible mechanisms will then be tied to specific neural structures and their interconnections. One conclusion will be that a given brain structure typically participates in multiple strategies and that performance on a recognition task also often engages disparate strategies and, hence, disparate sets of neural structures.

*1. Recognition as Part of Perception*

One possibility is to consider recognition as a part of perception. Arguably, recognition of simple features of a stimulus, or recognition that one stimulus differs from another, is really an aspect of perception. Perhaps we do not need to know anything about the world to recognize an emotion but are able to discriminate, categorize, and identify emotions solely on the basis of the geometric visual properties of a stimulus image. It is even conceivable (in principle) that such perceptual processing could be linked directly to language-related regions of the brain sufficient to produce the name of the emotion, in the absence of retrieving any other information associated with the stimulus (something akin to paired associate learning, for instance).

2. *Recognition via the Generation of Associated Knowledge*

However, recognition typically involves more than just perceptual information.Whenwe see a facial expression of fear, we can relate it not only to the percepts of other facial expressions in terms of its structure, but we can recognize that the person whose face we see is likely to scream, is likely to run away, has probably encountered something scary, and so on. None of that knowledge is present in the structure of the stimulus; it is present in our past experience with the world (and, to some limited extent, may even be present innately). A complex question concerns the precise mechanisms by which such knowledge might be retrieved. In general, the knowledge is not stored in any explicit format but rather relies on recipes for reconstructing knowledge by reactivation of the representations that were originally associated with one another when the knowledge was acquired (e.g., A. R. Damasio & Damasio, 1994). The simplest example of such a mechanism would be literal association, as when we see a face of fear and hear a scream at the same time and link the two henceforth in memory. In general, linking other knowledge with a perception of the facial expression will be vastly more complex and will rely on multiple higher order associations that may be fairly separated in time (e.g., seeing a face of fear and seeing the chasing tiger some time later), as well as on symbolic representations that, in humans, rely substantially on language (e.g., seeing a face of fear and merely being told that the person was afraid because he or she was running away from a tiger).

The general neural scheme for implementing the above mechanisms requires the binding of information between separate neural representations so that they can be processed as components of knowledge about the same concept. In the perceptual case, a stimulus activates multiple neural regions that represent particular aspects of its visual properties, and the coherent ensemble of these different bits of knowledge (the representations of the different properties of the stimulus) constitutes the perceptual mechanism that we discussed in Section 4 above. But, this mechanism can be extended beyond those neural regions that represent the visual properties of the stimulus to include those that represent knowledge not of the stimulus itself but of that with which it has been associated. The demand for integrating neural representations that are spatially separated in the brain would require extensive feedback connections as well as feedforward connections between different neural regions. One might thus envision a continuous dynamic interplay between feedforward, feedback, and horizontal information flow from which the brain constructs representations of visual stimuli (cf. Lamme, Super, & Spekreijse, 1998, for review). Schemes such as Ullman's (1995) counter streams or Edelman's (1987) re-entry both capture this idea: The representation of the stimulus itself, and of its associated knowledge, evolves contemporaneously such that the one continuously modulates the other and perception and recognition become parts of the same large-scale process.

*Recognition via the Generation of a Simulation*

The above mechanisms, although they rightly can be considered creative, are relatively direct: On linking together the various representations that give rise to components of the conceptual knowledge about the emotion that is signaled by the stimulus, the subject has available all the information necessary to recognize the emotion; all that is required to perform most recognition tasks now is an implementation of the reconstructed conceptual knowledge in terms of language so that the subject can tell us what he or she knows. But there are less direct routes that might come into play also. It may be that the explicit knowledge triggered in the above scheme is insufficient to recognize an emotion, perhaps because that particular emotion was never seen before or because the recipe for reconstructing knowledge about it provides insufficient detail. Another mechanism might attempt to generate conceptual knowledge using an inverse mapping that seeks to trigger those states normally antecedent to producing the facial expression. Such a mechanism would attempt to simulate in the observer the state of the person shown in the timulus by estimating the motor representations that gave rise to the observed stimulus. Once the observer has generated the state that the other person is presumed to share, a representation of this actual state in the observer could in turn trigger conceptual knowledge. Simulation thus still requires the triggering of conceptual knowledge, but the basis of the trigger is not a representation of someone else but rather a representation of ourselves (simulating the other person). The simulation hypothesis has recently received considerable attention due to experimental findings that appear to support it. In the premotor cortex of monkeys, Rizzolatti and colleagues have reported neurons that respond not only when the monkey prepares to perform an action itself but also when the monkey observes the same visually presented action performed by someone else (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Gallese & Goldman, 1999; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Various supportive findings have also been obtained in humans: Observing another's actions results in desynchronization in motor cortex as measured with Adolphs It thus appears that primates construct motor representations suited to performing the same action that they visually perceive someone else perform, in line with the simulation theory.

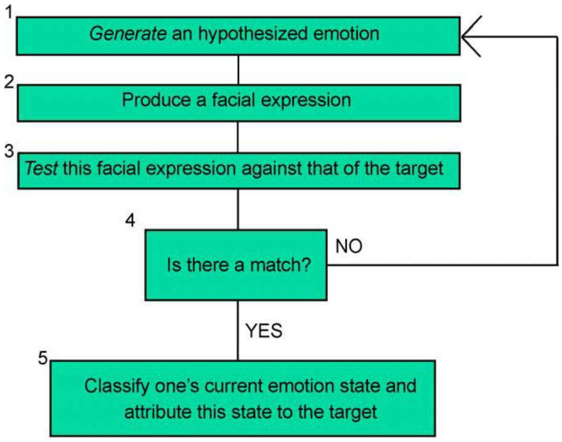
*B. The Development of Emotion Recognition*

The ability to discriminate and to recognize emotion from facial expressions develops in a complex fashion in infancy (Nelson, 1987; Saarni, Mumme, & Campos, 1997) and matures somewhat earlier in females than in males (see McClure, 2000, for a review).

Infants already orient to facelike stimuli at birth(Valenza, Simion, Macchi-Cassia, & Umilta, 1996), andthere is some evidence that this may depend primarilyon subcortical pathways, as indicated by the fact that they appear to process faces preferentially in temporal visual fields (Simion, Valenza, Umilta, & DallaBarba, 1998). Some basic emotions can be discriminated by 7-montholds (Nelson, Morse, & Leavitt, 1979; Soken & Pick, 1992), and responses in temporal visual cortices show some selectivity to the sight of faces in 2-month-old monkeys (Rodman, O Scalaidhe, & Gross, 1993). There is also evidence that Mechanism 6 above, recognition by simulation, may be engaged early on in life: Newborns already possess an innate ability to mimic some simple facial gestures (such as someone sticking out their tongue) (Meltzoff & Moore, 1983) that may be precursors to a more extensive ability to mimic and simulate others. Given the importance of communicating via facial expressions and other visual social signals, one would expect that infants who are born blind would be impaired in their social and emotional development. Although it has been exceedingly difficult to obtain unequivocal data on this issue, some studies do indeed suggest such an effect: Although even congenitally blind children express a range of facial emotions both spontaneously and volitionally, their expressions are not entirely normal (Cole, Jenkins, & Shott, 1989; Galati, Scherer, & Ricci-Bitti, 1997), and there is some suggestion that socioemotional development may be subtly abnormal as well (Troester & Brambring, 1992). In general, factors such as age and gender have not been investigated in detail for their contribution to differential performances in the experiments reviewed below. Although gender (Kesler-West et al., 2001) and age differences (Pine et al., 2001) in processing facial emotion do turn up in functional imaging studies, the evidence so far suggests that the effect sizes of these factors are relatively small compared to the effects of brain damage in lesion studies .

*C. Simulation Model*

The attributor starts by hypothesizing a certain emotion as the possible cause of the target's facial display and proceeds to enact that emotion, that is, produce a facsimile of it in her own system. She lets this facsimile (or pretend) emotion run its typical course, which includes the production of its natural facial expression, or at least a neural instruction to the facial musculature to construct the relevant expression. If the resulting facial expression, or the instruction to construct such an expression, matches the expression observed in the target, then the hypothesized emotion is confirmed and the attributor imputes that emotion to the target. The simulation interpretation of the paired-deficit findings would say that this is the sort of thing that happens in emotion interpreters who are normal with respect to the emotion in question.




Someone impaired in the relevant emotion area, however, cannot enact that emotion, or produce a facsimile of it. So she cannot generate the relevant facerelated downstream activity necessary to recognize the emotion. Hence, a recognition impairment specific to that emotion arises. Several issues about this model must be addressed. One question concerns the final phase of the postulated process, in which the system tries to match a constructed facial expression with the expression observed in the target. The representation of one's own facial expression is presumably a proprioceptive representation, whereas the representation of the target's expression is visual. How can one match the other? One possible answer is that the system has acquired an association between proprioceptive and visual representations of the same facial configuration, through some type of learning. Alternatively, there might be an innate cross-modal matching of the sort postulated by Meltzoff and Moore (1997) to account for neonate facial imitation. Second, there is a problem of how the generation process works. If candidate emotions are generated randomly, say, from the six basic emotions, the observer will have to covertly generate on average three facial expressions before hitting on the right one. This would be too slow to account for actual covert mimicry of displayed facial expressions, which occurs as early as 300 ms after stimulus onset (Dimberg & Thunberg, 1998; Lundquist & Dimberg, 1995). An alternative is to say that theoretical information is used to guide the generation process—though it isn't clear what theoretical information it would be. However, this proposal seems to turn the generate-and-test model into more of a theory–simulation hybrid rather than a pure simulationist model. Does this undercut the thrust of our simulationist argument? No. First, the simulational test phase of the generate-and-test heuristic is crucial, because without it the model cannot explain the paired deficits data. Second, the timing problems make this first model the least promising of the four we shall offer, and all of the other three are more purely simulationist in character.

## Conclusion

Paul Viola and Michael Jones presented an approach for object detection which minimizes computation time while achieving high detection accuracy. The approach was used to construct a face detection system which is approximately 15 faster than any previous approach. Preliminary experiments, which will be described elsewhere, show that highly efficient detectors for other objects, such as pedestrians, can also be constructed in this way.

New algorithms, representations, and insights where presented which are quite generic and may have broader application in computer vision and image processing.

The first contribution is a new a technique for computing a rich set of image features using the integral image. In order to achieve true scale invariance, almost all object detection systems must operate on multiple image scales. The integral image, by eliminating the need to compute a multi-scale image pyramid, reduces the initial image processing required for object detection significantly. In the domain of face detection the advantage is quite dramatic. Using the integral image, face detection is completed before an image pyramid can be computed.

## References

IEEE Paper on Classification and Recognition of Facial Expressions for Human Faces by Mrs. Sunanda Prabhakar Khandait.

Paper on Facial Expression Recognition'' by Mr. Rohit Dilip Gawade

IEEE Paper on Coding, Analysis, Interpretation, and Recognition of Facial Expressions by Ralph Gross, Iain Matthews, and Simon Baker.