Capturing Facial Actions in Video to Revive Expressions of Humans

KalaiSelvi R, Kavitha P, Shunmuganathan K L

Dept of Computer Science and Engineering, RMK Engineering College, Chennai, India.

Dept of Computer Science and Engineering, RMK Engineering College, Chennai, India.

Dept of Computer Science and Engineering, RMK Engineering College, Chennai, India.

Abstract— Emotion recognition in video is an interesting and important component in Human Machine Interaction (HMI) system. The recognition of emotional information is a key step toward giving computers the ability to interact more naturally and intelligently with people. Video-based facial expression recognition is a challenging problem in computer vision field. Audio-visual emotion recognition can be carried out with video sequence. The video sequence is a mixture of both audio and video information. This paper dealing only with the video information. The video sequence is segmented in to different frames. From that the target frame is selected and face detection is performed. The Facial Feature points around each facial component capture the detailed face shape information using Active Appearance Model. Action Unit classification represent the specific set of facial muscles. This Action Unit is compared with database AUs which are commonly used to describe the human emotion states. This paper introduces a framework based on Dynamic Bayesian Network (DBN) to represent facial evolvement in different levels. General experiments are performed to demonstrate the feasibility and success of the proposed model.

Keywords— Video segmentation, Facial Feature points, Active Appearance Model, Action Unit classification, Dynamic Bayesian Network.

I. INTRODUCTION

The revival of facial activities in video sequence is an important and challenging problem. Nowadays, plenty of computer vision techniques have been developed to track and recognize facial activities. Human computer interaction is an emerging field in computer science. To make the computer to be intelligent it must interact with human the way human and human interact. Mostly human interact through speech along with that they interact through physical gestures and postures which mainly include facial expressions.

Emotions are response or feeling to particular situation or environment. Emotions are an integral part of existence, as one smile to show greeting, glare when confused, raises voice when angry. This is because we understand other emotions and react based on that expression only through interactions. Computers are "emotionally challenged". Emotion may be characterized with respect to four common attributes: intensity, brevity, partiality and instability. Emotions usually are of relatively great intensity. Emotions usually express personal and interested perspectives found on a narrow target, such as a person or an object.

A facial expression is a key mechanism for understanding and conveying emotion. Studies have shown that interpreting facial expressions can significantly alter the interpretation of what is spoken as well as control the flow of a conversation. For ideal Human Computer Interfaces (HCI), we would desire that machines have this capability to interact with humans. Computer applications could better communicate by changing responses according to the emotional state of human users in various interactions. The machines that can understand emotion, we enhance the communication that exists between humans and computers. This would open a variety of possibilities in robotics and human-computer interfaces.

In this paper, six primary emotions commonly used even still that are said to be universal across human cultures: happiness, sadness, feat, disgust, surprise and anger.

Facial expression recognition is concerned with the recognition of certain facial movements about the underlying emotional state of the humans. The explanation for this comes from the relationship between

facial expression and underlying emotional states which do not necessarily map deterministically. The facial activity recognition system consists of two models: offline activity model, online activity model. Offline activity model uses training data and subjective domain knowledge. Online model uses techniques for facial feature points tracking and to get the measurements of facial motions.

Emotional Intelligence (EI) is a new discipline of knowledge. Philosophically, it refers to the competence to identify and express emotions, understand emotions, assimilate emotions in thought and regulate emotions in the self and in others. In the last decade, emotional intelligence has earned widespread publicity because of significant progress in experimental psychology. The most promising best-selling title on Emotional Intelligence is due to Daniel Goleman. According to Goleman, emotional intelligence was believed to have significant impact on individuals from the point of view of cognitive ability.

In Fig.1 different types of Emotion are shown. Emotions are the response of the human mind to external stimulus. Naturally, arousal of specific emotions depends on the involved phenomena in the external stimulus. Usually aroused emotions take some time for dispersal. Consequently, significant changes in the stimulus may cause arousal of different emotion with a trace of the previously aroused emotion in the psychological mind. These emotions are the basis for human-computer interaction.

| Happy | Fear | Sad |

| Normal | Anger | Surprise |

Fig. 1 Types of Emotion

II. LITERATURE SURVEY

The facial expression origins at 19th century, in that Darwin proposed the concept of Universal facial expressions in humans. In early 1970s, Ekman and Friesen performed studies of human facial expressions, to support this universality theory. Pantic [12] identify three basic problems of facial expression analysis approach. The problems are face detection in a facial image, facial expression data extraction and facial expression recognition. Mostly previous systems assume presence of a full frontal face view in the image or the image sequence is analyzed, generating some knowledge of the global face location.

Generally facial feature tracking classified into two types: Model free and Model based method. Valstar and Pantic [7] proposed Model free Method. Model free

approaches are general purpose point trackers without the prior knowledge of the object. Each feature point is usually detected and tracked individually by performing a local search for the best matching position. However, the model free methods are susceptible to the inevitable tracking errors due to the aperture problem, noise, and occlusion. Rogers et al [21] Model based method Active Shape Model (ASM) focus on explicitly modeling the shape of the objects. The ASM is subsequently proposed to combine constraints of both shape variation and texture variation. However, the discrete states still cannot describe the details of each facial component movement.

Current methods of expression recognition can be grouped into two categories: Image based and model based methods. Linda and Chandrapati [14] proposed the image based method Neural Network as a Useful Tool for Real-Time Facial Expression Recognition, which is used for extraction of facial features and expression /Action Units. It capable of recognizing facial expressions in real time with a relative accuracy in recognition, while leaving more computational capacity to process for emotion recognition. These kinds of networks are more appropriate for applications where each set of inputs has a different solution and is not linked to the results obtained in the previous iteration. Barlett et al [8] presented Support Vector Machines for recognizing facial expression. It acts as an active determinant for different expressions. It is trained with data sets created and different kernel models on database to increase accuracy of emotion recognition. But it tend to recognize each Action Units or certain Action Units combinations individually and statically directly from the image data, ignoring the semantic and dynamic relationships among AUs, although some of them analyze the temporal properties of facial features.

T.Ahonen *et al* [19] proposed a new facial representation strategy for still images based on Local Binary Pattern (LBP). The basic idea for developing the LBP operator was that two-dimensional surface textures can be identified by two complementary measures: 2D local spatial patterns and the gray scale difference. The histogram of these patterns for a local block of an image represents a local feature for the block. The histograms for all blocks can be concatenated to represent the feature vector for the image.

Facial Expression Recognition and Analysis challenge (FERA 2011) [6] consisted of the recognition of discrete emotion and detection of AUs with the ISIR laboratory. Even if the results were encouraging, the recognition rates remained low: the person-independent discrete emotion recognition did not exceed 75.2% (although the person-specific performance was 100%) and the AU detection only reached 62%.

Lien et al [4] employed model based approach as a set of Hidden Markov Models (HMMs) to represent the facial actions evolution in time. The classification is performed by choosing the Action Units or Action Units combination that maximizes the likelihood of the extracted facial features generated by the associated HMM. But it provides temporal dependencies among Action Units.

III. ARCHITECTURES

A. Existing Architecture

The overview of existing architecture is depicted in Fig. 2.The existing system uses image sequences, from that emotion is recognized. The existing method for facial feature tracking is Model free approaches. This method is susceptible to the inevitable tracking errors due to the aperture problem, noise, and occlusion. For expression recognition image based methods are used. This method is tending to recognize each AU or certain AU combinations individually and statically directly from the image data, ignoring the semantic and dynamic relationships among AUs.
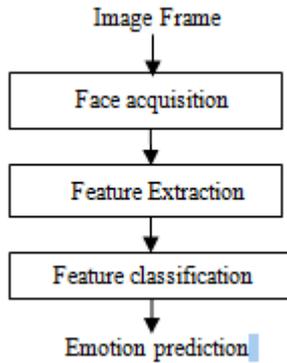

Fig.2 Emotion Recognition in Existing System

B. Proposed Architecture

The previous work demerits are overcome in proposed architecture. The overview of proposed system is shown in Fig. 3. In this emotion recognition is done in video sequences. The preprocessing steps face detection, Eye detection, Gabor transform are included, in which accuracy is increased. For Facial feature tracking, Model based approaches "Active Appearance model" is used, where shapes are represented by a set of feature points. Then the face is classified into different Action Units. The system is trained with prior knowledge of different Action Units dataset. This Action Units are compared to dataset Action Units. The dynamic relationships between Action Units are used for Expression Recognition. The existing weakness for emotion recognition is overcome using model based methods by making use of relationship among action Units and recognize the Action Units simultaneously.
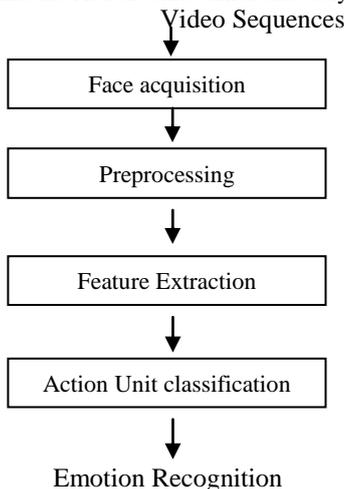

Fig. 3 Emotion Recognition in Proposed System

The Fig. 4 represents the overall flow of the proposed system. In this, a video sequence is given as input. From the video sequences frames are extracted. Then preprocessing steps are performed. Facial feature tracking is usually detect and track each feature points individually by performing a local search for the best matching position. Then Action Units classification is done and facial expression recognition systems usually try to recognize the emotion.
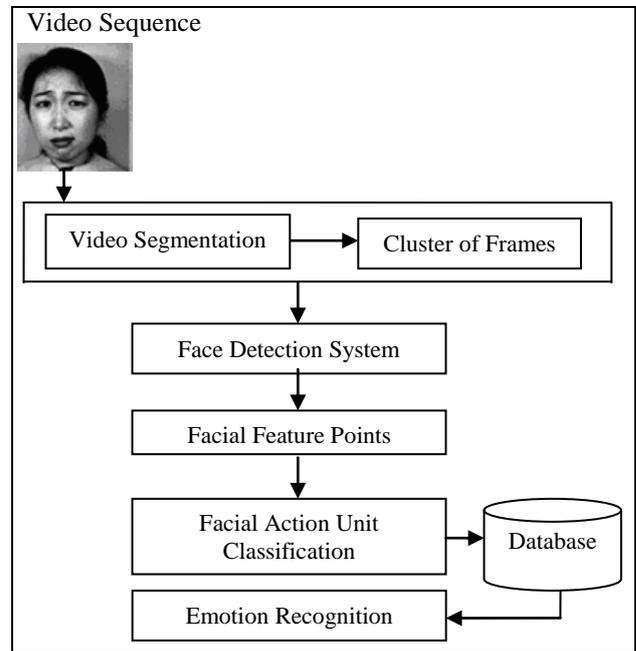

Fig. 4 Overall Architecture flow diagram

1) *Video Segmentation*: Video segmentation aims to partition the video into basic image sequences termed scenes and shots. A shot is defined as a set of successive frames taken without interval. A scene, on the other hand, is defined as the basic story-telling unit of the video. In video segmentation, the given video sequence is segmented into frames. These frames are stored in a specific location. Later the key frame is selected from stored location.

2) *Preprocessing*: Face detection is used for face localization which is a good stepping stone in facial expression. It is responsible for locating a face image within the input image and for determining the position of facial landmarks. Gabor Transform is extracted to represent the facial gestures or facial movements which are depicted in Fig.5 (a). Edge detector is used after Gabor transform, because it is very sensitive to soft edges and thus picks up a lot of edges using Canny Algorithm which is shown in Fig.5 (b).
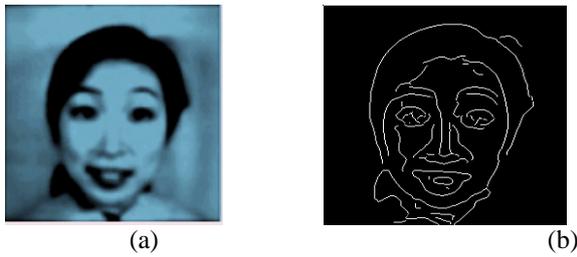
(a)   (b)

Fig. 5 Preprocessing steps. (a) Gabor Transform, (b) Edge Detection

3) *Facial Feature Tracking*: Facial feature tracking used to track facial feature points to find accurate location of face and face shape deformation. Accurate location and tracking of facial feature points are important in the applications such as animation, computer graphics, etc. Feature points are searched individually to analyze the models of shape variation so that the object shape can only deform in specific ways found in the training data. For facial feature tracking Active Appearance Model is used. Active Appearance model improve the robustness and accuracy of feature points search. In Active Appearance Model, the feature points positions are updated simultaneously, which indicates that the interactions within feature points are interdependent as illustrated in Fig.6.
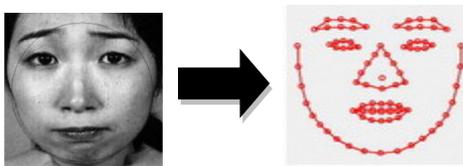

Fig. 6 Facial Feature points

4) *Action Unit Classification*: Action units are considered to be the smallest visually distinct facial movements. Action Units classification applies some form of pattern analysis technology to make classification decisions about the presence or absence of individual AUs in the input image. The global facial expression is to produce certain Action Units configurations, which in turn cause local muscle movements, and hence feature points movements. We infer expressions directly from the corresponding Action Units. The Dataset for Action Units Classification as shown in Table I. Only a set of common Action Units or Action Unit combinations, which produce significant facial actions.

TABLE I
DATASET FOR ACTION UNITS CLASSIFICATION

AU1	AU2	AU4	AU5	AU6
Inner brow raiser	Outer brow raiser	Brow Lowerer	Upper lid raiser	Cheek raiser

AU7	AU9	AU12	AU15	AU17
Lid Tighten	Nose Wrinkle	Lip corner puller	Lip corner depressor	Chin raiser

AU23	AU24	AU25	AU27
Lip tighten	Lip presser	Lips part	Mouth Stretch

5) *Emotion Recognition*: Expression Recognition maps the set of detected Action Units to one of the expressions which is of interest in the particular application. The selection of Action Units to be recognized is mainly based on the Action Units occurrence frequency. The co-occurrence and mutual exclusion relationships among AUs are achieved significantly by using Dynamic Bayesian Network for improvement of Action Unit recognition. Different action units both individually and in combination with other action units to recognize emotions. The Fig. 7 shows the different combination of Action Units to describe the Expression as Sad. Table II illustrates the AUs combination for six expressions.
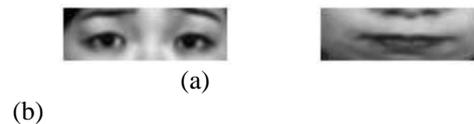

(a)
(b)

Fig. 7 Different Combinations of Action Units. (a) Outer brow raiser.
(b) Chin raiser. (AU2+AU17 enhances sad expression)

TABLE II
ACTION UNITS COMBINATIONS FOR DIFFERENT EXPRESSION

Emotion	Corresponding Action Units
Sad	AU2,AU17
Happy	AU6,AU12
Surprise	AU1,AU5,AU27,AU25
Anger	AU9,AU4,AU23
Fear	AU1,AU7,AU4
Normal	AU25,AU7

IV. APPLICATIONS

Facial emotion recognition has applications in: Facial emotion recognition has applications in medicine in treatment of Asperger. Physical clumsiness and a peculiar use of language are frequently reported. So it is easy to recognize the facial expression than languages. It also has application in video games. Most important use of any facial emotion technique is human-computer interaction to make intelligent tutoring systems. Affective computing is the study and development of systems and devices that can recognize, construe, process, and simulate human affects. The machine should interpret the emotional state

of humans and adapt its behavior to them, giving an suitable response for those emotions. Face Expression tracking is to drive real time avatar based chat systems. Face recognition technology for law enforcement applications. It is also used in Robotics.

V. CONCLUSIONS

In this paper, we proposed an approach to recognize facial emotion in the video. Incorporating emotive information in computer-human interfaces will allow for much more natural and efficient interaction paradigms. The previous work was done in image technology. The proposed model overcomes the previous work by recognizing in video technology. This paper improves the emotion recognition through preprocessing steps, facial feature tracking and Action Unit classification. We evaluated our system in terms of accuracy for a variety of interaction scenarios and found the results for controlled experiments to compare favorably to previous approaches to expression recognition. A current weakness in this area of facial study is still the lack of comparable databases. We would also like to encourage the creation and use of common data sets in this area as a means to strengthen comparison and fine-tuning of techniques.

ACKNOWLEDGMENT

REFERENCES

[1] Cassio P. de Campos and Q. Ji, "Efficient structure learning of Bayesian networks using constraints," *J. Mach. Learn. Res.,*vol. 12, pp. 663–689, Mar. 2011.

[2] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis.Comput.,* vol. 27, no. 6, pp. 803–816, 2009.

[3] D.Heckerman, D. Geiger, and D.M. Chickering "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learn.,* vol. 20, no. 3, pp. 197–243, 1995.

[4] J. J. Lien, T.Kanade, J. F. Cohn, and C. Li, "Detection, tracking, and classification of action units in facial expression," *J. Robot. Auto. Syst.,*vol. 31, no. 3, pp. 131–146, 2011..

[5] J. Chen and Q. Ji, "A hierarchical framework for simultaneous facial activity tracking," in Proc. IEEE *Int. Conf. Autom. Face Gesture Recognit.,*Mar. 2011, pp. 679–686.

[6] M.F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer. The first facial expression recognition and analysis challenge. In Automatic Face & Gesture Recognition and Workshops, 2011 IEEE International Conference on, 2011.

[7] M.Valstar and M.Pantic, "Fully automatic recognition of the temporal phases of facial actions," IEEE *Trans. Syst., Man, Cybern. B, Cybern.,* vol. 42, no. 1, pp. 28–43, Feb. 2012.

[8] M.S.Bartlett, G.Littlewort, M.G.Frank, C.Lainscsek, I.Fasel, and J.R.Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," in Proc. *IEEE Conf. Compute. Vis. Pattern Recognition.,* vol. 2. Jun. 2005.

[9] Y. Cheon and D. Kim. Natural facial expression recognition using differential-AAM and manifold learning. Pattern Recognition, 2009.

[10] Y. Tong, Y. Wang, Z. Zhu, and Q. Ji, "Robust facial feature tracking under varying face pose and facial expression," Pattern Recognition., vol. 40, no. 11, pp. 3195–3208, 2007.

[11] Z.Zeng, M. Pantic, G. I. Roisman, and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 2009.

[12] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. I*EEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[13] T. F. Cootes, G. J. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685,Jun. 2001.

[14] Linda O., Chandrapati S., Tokuhiro A., "Neural Network as a Useful Tool for Real-Time Facial Expression Recognition", *Proceedings of 17th Conference on Artificial Neural Networks in Engineering, St Louis ,USA, Nov 2007.*

[15] Srivastava, R.; Roy, S.; Shuicheng Yan; Sim, T., "Accumulated motion images for facial expression recognition in videos," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* , vol., no., pp.903,908, 21-25 March 2011.

[16] Datcu, D.; Rothkrantz, L. J M, "Automatic recognition of facial expressions using Bayesian belief networks," *Systems, Man and Cybernetics, 2004 IEEE International Conference on* , vol.3, no., pp.2209,2214 vol.3, 10-13 Oct. 2004.

[17] Cruz, A.; Bhanu, B., "A biologically inspired approach for fusing facial expression and appearance for emotion recognition," *Image Processing (ICIP), 2012 19th IEEE International Conference on* , vol., no., pp.2625,2628, Sept. 30 2012-Oct. 3 2012.

[18] Mikels, J.A. et al. (2005). Emotional category data on images from the International Affective Picture System, Behav. Res. Methods 37, 626–630.

[19] Ahonen, T.; Hadid, A. &Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI, 28, pp. 2037-2041, December 2006.

[20] T. Wu, N. Butko, P. Ruvolo, J. Whitehill, M. Bartlett, and J. Movellan, "Action unit recognition transfer across datasets", in Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, Mar 2011, pp. 889 v896.

[21] M. Rogers and J. Graham, "Robust active shape model search," in *Proc.Eur. Conf. Comput. Vis.*, 2002, pp. 517–530.