Int. Conf. on Signal, Image Processing Communication & Automation, ICSIPCA

Man Machine Interface –A Review

Komala K[1] and B Jayadevappa[2]
[1]Asst Professor, Department of Medical Electronics, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India
komala1724@gmail.com
[2]Professor, Department of Electronics and Instrumentation Engineering, JSSATE, Bengaluru, Karnataka, India
jayadevappab@jssateb.ac.in

***Abstract*— The way a human communicate with any intelligent and computing system has a great history which dates back to the advent of a computer system. The first and the foremost type of communication was by using a machine understandable languge which was coded using binary numbers consisting of 1s and 0s. From then, there has been a lot of research and improvement in the way a human communicate with the computer and intelligent systems. This paper reviews one such research area which concentrates on the process of transreceiving the message between the human and the computer system by using just different types of expressions and different expressive features human beings exhibit during the though process which takes place with in him/her. The paper extensively reviews the process flow which may be considered as a preliminary stage of decision making by an intelligent system which sends the probable command delivered by the user to the computer system.**

***Index Terms*— Human computer interaction, Intelligent systems, Artificial learning.**

I. INTRODUCTION

Over the past half-century, there have been several important Developments spanning multiple approaches for qualifying and quantifying emotions related to human behavior.

Since the early studies of human behavior, emotion has attracted the interest of researchers in many disciplines of Neurosciences and Psychology. More recently, it is a growing field of research in computer science and machine learning. The challenge is to exploring how the expression of emotion is perceived by listeners and how to represent and automatically detect a subject's emotional state by speech ,facial expressions ,body gestures, eye gaze,,, cognitive modelling etc.

In contrast with most previous studies, conducted on artificial data with archetypal emotions. Human-Computer Interaction (HCI) aims at significant interaction between human and machines. Facial expressions can be used as an efficient way of emotion detection, thus facilitating HCI.A natural way of interaction between man and machine can be obtained by detection and classification of emotional (facial) expressions.

However emotional expressions and its intensity vary from person to person, even the expression of the same subject (person) varies with respect to time. Hence designing a generalized method for emotional expression analysis is a very tough task. However, specified exiting methods and algorithms can be employed to analyze emotional expressions of a Single person accurately.

After reviewing, studying and evaluating the theories and methods employed in the field of emotion research, it was observed with varying strengths and weaknesses in all of them. The field of emotion research is young

Grenze ID: 02.ICSIPCA.2017.1.8

And rapidly growing. In our opinion, the reviewed methods had a similar level of exposure and validation. Emotions play an essential role in social interactions and facilitate rational decision making and perception. The human computer interaction have started their investigation and tried to understand different causes and effects. Some progress has been made in developing ‘‘affective systems” that are capable of recognizing and appropriately responding to human emotions, and ultimately making human–computer interaction experiences more effective and pleasurable.

Here the approach has been to present the existing methodologies and to explore the advantages and disadvantages of individual traits for emotion recognition task , against their specific research objectives it is needed to develop a robust emotion recognition system using integration of multiple cues from the source of human information in order to achieve high accuracy of the system. A combination of low-level features, high-level reasoning, and natural language processing is likely to provide the best emotion inference for the emotion recognition system. Further the paper reviews extensively the emotion recognition techniques available in literature which is followed by the brief presentation of the authors' preliminary idea of having a human computer interface based on different emotions of the human beings. Further, session four discusses the proposed application areas such a system. The paper has been concluded in section 5 with a brief note about the future work planned in this direction.

II. EXISTING LITERATURE

Siamak Yousefi et al [1] combines empirical spiking data from primary motor cortex and mathematical analysis to explain this loss in performance. We focus on angular velocity as a meaningful and convenient measure of smoothness. Our results demonstrate that angular velocity in the trajectory is approximately proportional to change in target probability. The constant of proportionality equals the difference in heading between parallel filters from the two most probable targets, suggesting a smoothness benefit to more narrowly spaced targets. Simulation confirms that measures to smooth the data likelihood also improve the smoothness of hybrid trajectories, including increased ensemble size and uniformity in preferred directions.

Jun Yu and Zeng-Fu Wang [2] proposed a multiple inputs-driven realistic facial animation system based on 3-D virtual head for human–machine interface is proposed. The system can be driven independently by video, text, and speech, thus can interact with humans through diverse interfaces. The combination of parameterized model and muscular model is used to obtain a tradeoff between computational efficiency and high realism of 3-D facial animation. The online appearance model is used to track 3-D facial motion from video in the framework of particle filtering, and multiple measurements, i.e., pixel color value of input image and Gabor wavelet coefficient of illumination ratio image, are infused to reduce the influence of lighting and person dependence for the construction of online appearance model. The tri-phone model is used to reduce the computational consumption of visual co-articulation in speech synchronized viseme synthesis without sacrificing any performance. The objective and subjective experiments show that the system is suitable for human–machine interaction.

J. F. Wu et al., [3] proposed system uses a consumer wireless recording device to collect EOG and employs new encoding/decoding paradigms to convey users' intentions with EOG from eye movements including blinking and looking up. The simplicity and mobility of the system provides a comfortable and practical solution to HMI. Furthermore, to reduce the hardware complexity and power consumption of the signal processing modules of the EOG-based HMI system, a novel multiplierless implementation is developed, where all the algorithms involved, such as bandpass filtering, wavelet filtering, and support vector machine, can be realized using a limited number of adders and shifters only. Experimental results show that the proposed system offers a simple, practical, and yet reliable EOG-based HMI with low complexity and power consumption.

Siamak Yousefi et al [4] combines empirical spiking data from primary motor cortex and mathematical analysis to explain this loss in performance. We focus on angular velocity as a meaningful and convenient measure of smoothness. Our results demonstrate that angular velocity in the trajectory is approximately proportional to change in target probability. The constant of proportionality equals the difference in heading between parallel filters from the two most probable targets, suggesting a smoothness benefit to more narrowly spaced targets. Simulation confirms that measures to smooth the data likelihood also improve the smoothness of hybrid trajectories, including increased ensemble size and uniformity in preferred directions.

Jun Yu and Zeng-Fu Wang [5] proposed a multiple inputs-driven realistic facial animation system based on 3-D virtual head for human–machine interface is proposed. The system can be driven independently by

video, text, and speech, thus can interact with humans through diverse interfaces. The combination of parameterized model and muscular model is used to obtain a tradeoff between computational efficiency and high realism of 3-D facial animation. The online appearance model is used to track 3-D facial motion from video in the framework of particle filtering, and multiple measurements, i.e., pixel color value of input image and Gabor wavelet coefficient of illumination ratio image, are infused to reduce the influence of lighting and person dependence for the construction of online appearance model. The tri-phone model is used to reduce the computational consumption of visual co-articulation in speech synchronized viseme synthesis without sacrificing any performance. The objective and subjective experiments show that the system is suitable for human–machine interaction.

J. F. Wu et al., [6] proposed system uses a consumer wireless recording device to collect EOG and employs new encoding/decoding paradigms to convey users' intentions with EOG from eye movements including blinking and looking up. The simplicity and mobility of the system provides a comfortable and practical solution to HMI. Furthermore, to reduce the hardware complexity and power consumption of the signal processing modules of the EOG-based HMI system, a novel multiplier less implementation is developed, where all the algorithms involved, such as band pass filtering, wavelet filtering, and support vector machine, can be realized using a limited number of adders and shifters only. Experimental results show that the proposed system offers a simple, practical, and yet reliable EOG-based HMI with low complexity and power consumption.

III. PROPOSED SYSTEM

A typical issue of multimodal data processing so far is that the multisensory data are typically processed separately and only combined at the end. Yet this is almost certainly incorrect; people display audio and visual communicative signals in a complementary and redundant manner. A multimodal system should be able to deal with these imperfect data and generate its conclusion so that the certainty associated with it varies in accordance to the input data. A way of achieving this is to consider the time-instance versus time-scale dimension of human nonverbal communicative signals.

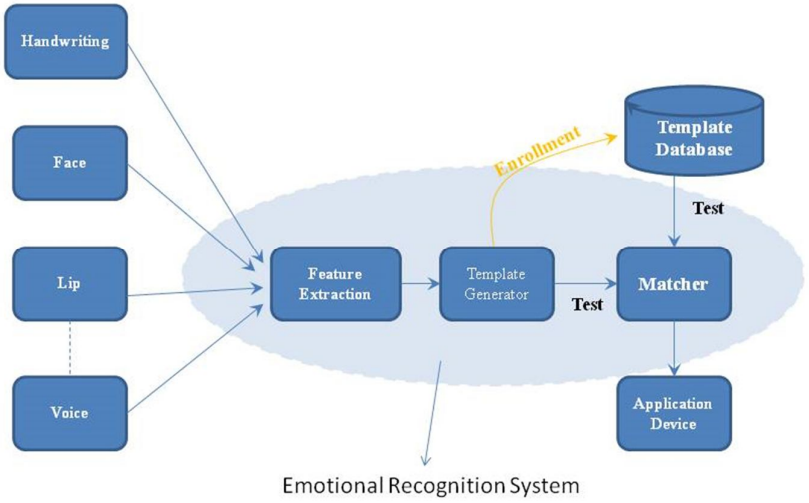

Figure 1. Block Diagram of the Proposed Emotion Recognition System

As shown in the block diagram, there are many features of the human being which needs to be studied before jumping into a conclusion about what the human is trying to communicate. Such as Hand writing, facial expression, Lip position, its size, orientation etc and the voice as well. With all these details which has to be fed into the decision making system, the system will be able to extract the required feature by using the principles of image processing like enhancement, segmentation, object/ pattern recognition etc so that combining all these features a typical template will be generated. This template which will be matched with the database of huge size so as to jump into a conclusion about what the user is trying to communicate.

A typical example is as shown in figure 2 where different features of the facial expression has been taken into consideration where the position of eye lobes, nose orientation, orientation of eyes, lips, and also the position of the eye ball has been considered and experimented to have a conclusion that the human being is exhibiting the characteristic features corresponding to the respective expression.

In the same way different input signals as shown in the block diagram (figure 1) would have their own characteristic feature based on the different orientations and types of such signals. The first block, that is the input data of the human characteristic, hand writing, human being cannot have the same hand writing when he is enthusiastic and when he is not comfortable in writing. To evaluate the same, many parameters may be considered such as the spacing between each letter, the angles of different letters, the size of each letter, the spelling and grammatical mistakes etc, if typically analyzed, will definitely provide a clue about the characters the human exhibiting and thereby the typical things what is being communicated by him.

Here multiple signals are being analyzed and it is for sure that the system will take a lot of time to process the data and to give away the output about the subject which has to be communicated out by the human so that the same can be used as an input for different systems like a simple computer (where just the hand writing on a touch pad or the voice is sufficient) and also to a much complex systems like an intelligent system where many or all the said input signals are used, feature extracted and jumped to a conclusion about the input being fed in to it.

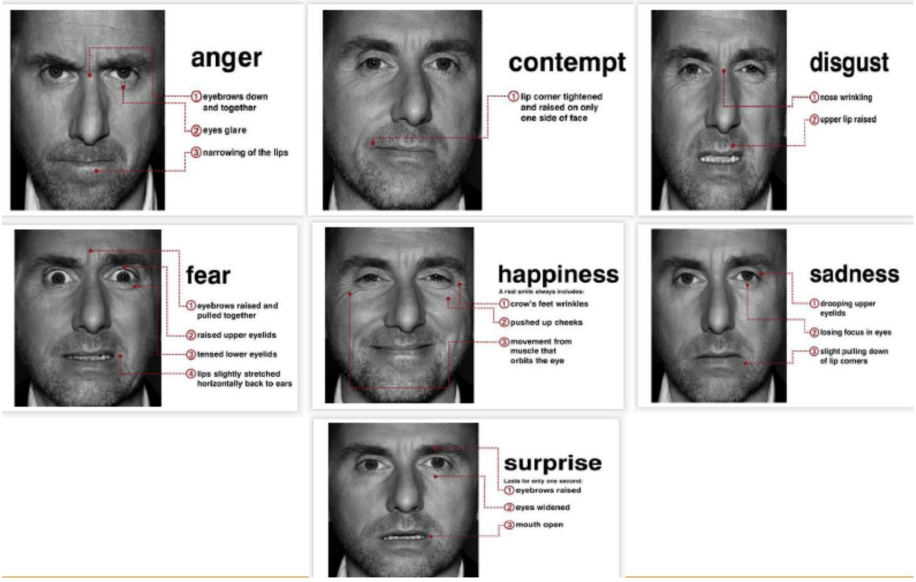

Figure 2. Different Facial Expressions

IV. CONCLUSION

Several Survey reports says that till date we could not achieve 100% accuracy in recognizing either a speaker or his emotion. When the emotional state of speaker differs in the testing phase the recognition rate decreases significantly. As a consequence, emotions, largely overlooked in early efforts to develop machine intelligence, are increasingly regarded as an area of important research. Emotion modulates almost all modes of human communication facial expression, gestures, posture, tone of voice, choosing of words, respiration, skin temperature and clamminess, etc. Emotions can significantly change the message: sometimes it is not what was said that is the most important, but how it was said. Faces tend to be the most visible form of emotion communication, but they are also most easily controlled in response to different social situations when compared to the voice and other ways of expression.

As per the literature survey emotion recognition is most likely to be accurate when it combines multiple cues or modalities, information about the user's context, situation, goal, and preferences. A combination of low-level features, high-level reasoning, and natural language processing is likely to provide the best emotion inference. Considering all these aspects, experts believe that multimodal context-sensitive human-computer

interaction is likely to become the single most widespread research topic of the Pattern and artificial intelligence research community. Although there were significant advances in the fields of video and audio processing, pattern recognition, computer vision, and affective computing, the realization of a robust, multimodal, adaptive, context-sensitive analyzer of human nonverbal affective state is far from being a reality. Currently, the researchers have to cope with the lack of a better understanding of individual- and context-dependent human behavior and with a better integration of multiple sensors and pertinent modalities according to the model of human sensory system.

Besides these problems there are other social and ethical issues that should be considered. The context-sensitive multimodal system that is supposed to interact with the human should not invade the user's privacy. By taking all of these aspects into account, we hope to be able to develop into the near future multimodal context-sensitive systems that are smart, perceptually aware, recognize the context in which they act, can adapt to their users, and can understand how they feel, and respond appropriately.

REFERENCES

[1] Siamak Yousefi, Alex Wein, Kevin C. Kowalski, Andrew G. Richardson, and Lakshminarayan Srinivasan, "Smoothness as a Failure Mode of Bayesian Mixture Models in Brain–Machine Interfaces," IEEE Transactions On Neural Systems And Rehabilitation Engineering, Vol. 23, No. 1, January 2015

[2] Jun Yu and Zeng-Fu Wang, "A Video, Text, and Speech-Driven Realistic 3-D Virtual Head for Human–Machine Interface," IEEE Transactions On Cybernetics, Vol. 45, No. 5, May 2015.

[3] J. F. Wu, A. M. S. Ang, K. M. Tsui, H. C. Wu, Y. S. Hung, Y. Hu, J. N. F. Mak, S. C. Chan, and Z. G. Zhang, "Efficient Implementation and Design of a New Single-Channel Electrooculography-Based Human–Machine Interface System," IEEE Transactions On Circuits And Systems, Vol. 62, No. 2, February 2015.

[4] Siamak Yousefi, Alex Wein, Kevin C. Kowalski, Andrew G. Richardson, and Lakshminarayan Srinivasan, "Smoothness as a Failure Mode of Bayesian Mixture Models in Brain–Machine Interfaces," IEEE Transactions On Neural Systems And Rehabilitation Engineering, Vol. 23, No. 1, January 2015.

[5] Jun Yu and Zeng-Fu Wang, "A Video, Text, and Speech-Driven Realistic 3-D Virtual Head for Human–Machine Interface," IEEE Transactions On Cybernetics, Vol. 45, No. 5, May 2015.

[6] J. F. Wu, A. M. S. Ang, K. M. Tsui, H. C. Wu, Y. S. Hung, Y. Hu, J. N. F. Mak, S. C. Chan, and Z. G. Zhang, "Efficient Implementation and Design of a New Single-Channel Electrooculography-Based Human–Machine Interface System," IEEE Transactions On Circuits And Systems, Vol. 62, No. 2, February 2015.

[7] Prof. Anil K. Jain, Dr. Arun A. Ross, Dr. Karthik Nandakumar, "Security Of Biometric Systems", Introduction to Biotmetrics 2011, pp 259-306 springier

[8] B. Reeves and C. Nass, The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places, Cambridge Univ. Press, 1996.

[9] R. W. Picard, Affective Computing, MIT Press, Cambridge, MA, 1997.

[10] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique," Psychological Bulletin 115(2), pp. 268–287, 1994.

[11] P. Ekman and W. Friesen, Facial Action Coding System: Investigator's Guide, Consulting Psychologists Press, 1978.

[12] K. Mase, "Recognition of facial expression from optical flow," IEICE Trans. E74(10), pp. 3474–3483, 1991.

[13] N. Ueki, S. Morishima, H. Yamada, and H. Harashima, "Expression analysis/synthesis system based on emotion space constructed by multilayered neural network," Systems and Computers in Japan 25(13),pp. 95–103, 1994.

[14] A.Lanitis, C. Taylor, and T. Cootes, "A unified approach to coding and interpreting face images," in Proc.International Conf. On Computer Vision, pp. 368–373, 1995.

[15] M. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametricodels of image motion," in Proc. International Conf. On Computer Vision, pp. 374–381, 1995.

[16] M. Rosenblum, Y. Yacoob, and L. Davis, "Human expression recognition from motion using a radial basis function network architecture," IEEE Trans. On Neural Network 7(5), pp. 1121–1138, 1996.