International Journal of Instrumentation Electrical Electronics Engineering
p- ISSN: 2321 – 0621 e- ISSN: 2321- 063X

Meta-Analysis of the First Facial Expression Recognition Challenge

* Vutukuri Samba Siva Rao [1], Dr.R.V.Krishnaiah[2]
[1]PG Student (M.Tech VLSI&ES), Dept. of ECE, DRK Institute of Science & Technology, Hyderabad, AP, India
[2] Professor, Dept. of ECE, DRK Institute of Science & Technology, Hyderabad, AP, India

Abstract: Automatic facial expression recognition has been an active topic in computer science for over two decades, in particular facial action coding system action unit (AU) detection and classification of a number of discrete emotion states from facial expressive imagery. Our system is designed by using ARM 32-bit micro controller which supports different features and algorithms for development of first facial recognition. The webcam combines video sensing, video processing and communication within a single device it captures a video stream like different expressions of face, computes the information and transfers the compressed video stream to the ARM micro controller. The image it received is processed by using image processing algorithms and processed image is classified by using PCA algorithms and identified expressions are displayed on display unit. Our system is designed by using S3C2440 micro controller developed by Samsung which was called as friendly ARM or mini 2440 board.

Keywords: ARM, PCA, S3C2440.

1. INTRODUCTION

Computers and other powerful electronic devices surround us in ever increasing numbers, with their ease of use continuously being improved by user-friendly interfaces.Yet, to completely remove all interaction barriers, the nextgeneration computing (a.k.a. pervasive computing, ambient intelligence, and human computing) will need to develop human-centered user interfaces that respond readily to naturally occurring multimodal human communication. An important functionality of these interfaces will be the capacity to perceive and understand the user's cognitive appraisals, action tendencies, and social intentions that are usually associated with emotional experience. Because facial behavior is believed to be an important source of such emotional and interpersonal information, automatic analysis of facial expressions is crucial to human–computer interaction.

Facial expression recognition, in particular facial action coding system (FACS) action unit (AU) detection and classification of facial expression imagery in a number of discrete emotion categories, has been an active topic in computer science for some time now, with arguably the first work on automatic facial expression recognition being published in 1973.Many promising approaches have been reported. The first survey of the field was published in 1992 and has been followed up by several others. However, the question remains as to whether the approaches proposed to date actually deliver what they promise. To help answer that question, we felt that it was time to take stock, in an objective manner, of how far the field has progressed.

2. GEMEP-FERA DATA SET

To be suitable to base a challenge on, a data set needs to satisfy two criteria. First, it must have the correct labeling, which in our case means frame-by-frame AU labels and event coding of discrete emotions. Second, the database cannot be publicly available at the time of the challenge. The GEMEP database is one of the few databases that meet both conditions and was therefore chosen for this challenge.

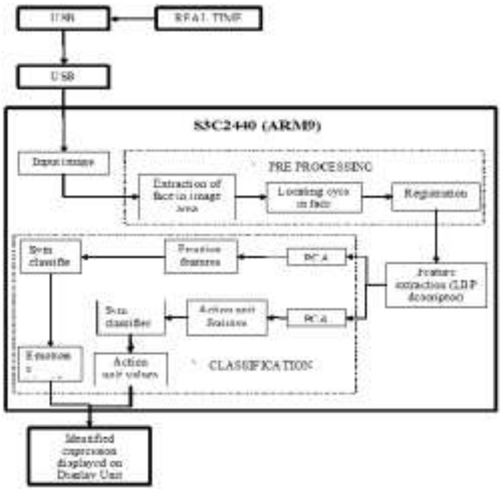

Figure 1 Block Diagram of The System

By no means does the GEMEP-FERA data set constitute the entire GEMEP corpus. In selecting videos from the GEMEP corpus to include in the GEMEP-FERA data set, the main criterion was the availability of a sufficient number of examples per unit of detection for training and testing. It was important that the examples selected for the training set were different from the examples selected for the test set.

PARTITIONING

For the AU detection subchallenge, we used a subset of the GEMEP corpus annotated with the FACS. The 12 most commonly observed AUs in the GEMEP corpus were selected (see Table I). To be able to objectively measure the performance of the competing facial expression recognition systems, we split the data set into a training set and a test set. A total of 158 portrayals (87 for training and 71 for testing) were selected for the AU subchallenge. All portrayals are recordings of actors speaking one of the two pseudolinguistic phoneme sequences. Consequently, AU detection is to be performed during speech. The training set included seven actors (three men), and the test set included six actors (three men), half of which were not present in the training set. Even though some

International Journal of Instrumentation Electrical Electronics Engineering
p- ISSN: 2321 – 0621 e- ISSN: 2321- 063X

actors were present in both training and test sets, the actual portrayals made by these actors were different in both sets.

For the emotion subchallenge, portrayals of five emotional states were retained: anger, fear, joy, sadness, and relief. Four of these five categories are part of what Ekman called basic emotions as they are believed to be expressed universally by specific patterns of facial expression. The fifth emotion, relief, was added to provide a balance between positive and negative emotions but also to add an emotion that is not typically included in previous studies on automatic emotion recognition. Emotion recognition systems are usually modeled on the basic emotions; hence, adding "relief" made the task more challenging.

CHALLENGE PROTOCOL

The challenge is divided into two subchallenges. The goal of the AU detection subchallenge is to identify in every frame of a video whether an AU was present or not (i.e., it is a multiplelabel binary classification problem at frame level). The goal of the emotion recognition subchallenge is to recognize which emotion was depicted in that video, out of five possible choices (i.e., it is a single-label multiclass problem at event level). The challenge protocol is divided into five stages. First, interested parties registered for the challenge and signed the EULA to gain access to the training data. Then, they trained their systems. In the third stage, the participants downloaded the test partition and generated the predictions for the subchallenges they were interested in. They then sent their results to the FERA 2011 organizers who calculate their scores. In the case of the FERA 2011 challenge, the participants then submitted a paper describing their approach and reporting their scores to the FERA 2011 workshop. Researchers who intend to follow this benchmark protocol after the FERA 2011 challenge are assumed to submit a paper to another relevant outlet.

Because of concerns regarding the ease with which the emotion labels can be guessed from the video data, the organizers introduced a secondary test for the emotion subchallenge held the day before the FERA 2011 workshop. The secondary test set contained 50 previously unreleased GEMEP videos displaying one of the five discrete emotions used in the challenge. Participants had the choice to either send their end-to-end programs to the organizers, who then run the secondary test for them, or they could choose to perform the test on their own hardware on-site the day before the workshop. The scores for this secondary test set were not to influence the participant ranking in the emotion detection subchallenge, but they were announced during the FERA 2011 workshop, on the FERA 2011 website, and in this paper. All participants but one performed this secondary test.

3. COMPETITION RESULTS

The number of parties who showed interest in participating in the FERA 2011 challenge indicates that the facial expression analysis field is of a moderate size. The challenge data were downloaded by 20 teams, of which 15 participated in the challenge and submitted a paper to the FERA 2011 workshop. Of the 15 papers, 11 papers were accepted for publication, based on a double-blind peer review process.

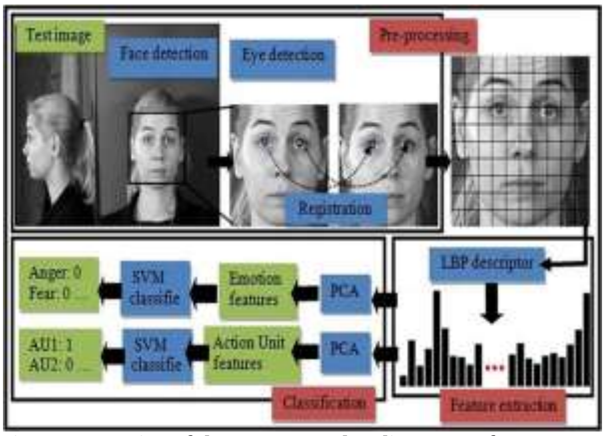

Figure 2 Overview of the FERA 2011 baseline system for detection of 12 AUs

In total, ten teams participated in the emotion recognition subchallenge, and five teams took part in the AU detection subchallenge (three teams participated in both subchallenges).

Table I shows the scores attained in the emotion recognition subchallenge. As can be seen, nine out of ten participating systems outperform the baseline approach on the full test set. The winning team, Yang and Bhanu of the University of California Riverside, attained an overall 83.8% classification result. It is interesting to note the person-specific results obtained by the multi-institute team of the University of Illinois, Urbana–Champaign, and the University of Missouri, Columbia. The proposed method, which included an automatic face recognition module, attained a perfect emotion recognition score on the subject-dependent test set.

Table 1 Average classification rates over all emotions for the emotion recognition subchallenge and average f1-measure

Participant	AU detection			Emotion detection			
	Person-independent	Person-specific	Overall	Person-independent	Person-specific	Overall	Secondary
ANU [14]	N.A.	N.A.	N.A.	0.649	0.838	0.734	0.700
ISIR [48]	**0.633**	**0.576**	**0.620**	N.A.	N.A.	N.A.	N.A.
KIT [21]	0.543	0.473	0.523	0.658	0.944	0.773	0.760
MIT-Cambridge [5]	0.470	0.422	0.461	0.448	0.433	0.440	0.480
Montreal [12]	N.A.	N.A.	N.A.	0.579	0.870	0.700	**0.96**
NUS [53]	N.A.	N.A.	N.A.	0.636	0.730	0.672	0.640
Riverside [66]	N.A.	N.A.	N.A.	**0.752**	0.962	**0.838**	0.860
QUT [8]	0.530	0.460	0.510	0.624	0.554	0.600	0.00
UCLIC [34]	N.A.	N.A.	N.A.	0.609	0.837	0.700	0.740
UCSD1 [30]	N.A.	N.A.	N.A.	0.714	0.837	0.761	0.640
UCSD2 [65]	0.604	0.539	0.583	N.A.	N.A.	N.A.	N.A.
UIUC-UMC [55]	N.A.	N.A.	N.A.	0.655	**1.00**	0.798	0.780
Baseline	0.453	0.423	0.451	0.440	0.730	0.560	N.A.

The secondary on-site emotion recognition test was introduced to perform a sanity check regarding the reported results. That is, it was used to ensure that nobody had either grossly inflated their performance results by guessing the emotion labels of the original test set or had in fact relied on some form of manual processing of the data. Participants were allowed to apply bug fixes to their original entry, which in at least one case led to a significant improvement in results.

AU Detection

The results for the AU detection subchallenge are shown per partition in Table I, and overall results per AU for each team are shown in Table II. The winner of the AU detection subchallenge was the team of Senechal et al., from the Institut des Systemes Intelligents et de Robotique, Paris. Their method attained an F1-measure of 63.3%, averaged over all 12 AUs. This is well above the baseline's 45.3% but still very far off from a perfect AU recognition.

Looking at individual AUs, we can see that AU1, AU2, AU6, and AU12 are consistently detected well by all participants, while AU4, AU5, AU10, AU17, AU18, and AU26 were consistently detected with low accuracy. AU25, parting of the lips, is detected with high accuracy by all participants except QUT. The authors noted that this may have been due to an inability to deal with speech effectively.

Contrary to what would normally be expected, Table I shows that performance on the person-specific partition was consistently worse than on the person-independent part. Unfortunately, given the relatively small size of the test partition, this is probably simply because the videos selected for the personspecific part may have been that much more challenging than those included in the person-independent part.

Table 2 f1-measures per au for every participant in the au detection

AU	ISIR	KIT	MIT-Camb.	QUT	UCSD	Avg
1	**0.809**	0.606	0.681	0.780	0.634	0.702
2	**0.731**	0.520	0.635	0.723	0.636	0.649
4	0.582	0.529	0.446	0.433	**0.602**	0.518
6	**0.833**	0.822	0.739	0.658	0.759	0.762
7	**0.702**	0.554	0.323	0.553	0.604	0.547
10	0.475	0.467	0.328	0.468	**0.565**	0.460
12	0.803	0.798	0.658	0.778	**0.832**	0.774
15	**0.245**	0.065	0.114	0.156	0.193	0.155
17	**0.557**	0.518	0.300	0.471	0.499	0.469
18	0.431	0.329	0.127	**0.448**	0.345	0.336
25	**0.850**	0.800	0.815	0.311	0.815	0.718
26	**0.576**	0.515	0.475	0.537	0.515	0.524

4. CONCLUSIONS

The paper "Meta-Analysis of the First Facial Expression Recognition Challenge" has been successfully designed and tested. It has been developed by integrating features of all the hardware components and software used. Presence of every module has been reasoned out and placed carefully thus contributing to the best working of the unit. Secondly, using highly advanced ARM9 board and with the help of growing technology the project has been successfully implemented.

Another issue that arose during the challenge is the choice of performance measure. It is well known that, in a heavily unbalanced data, such as that of the AU detection subchallenge, the classification rate is not a suitable measure. A naive classifier based on the prior probability of the classes will give an overoptimistic representation of the problem and is very likely to outperform systems that try to detect both classes with equal priority. The detection of AUs, however, is still far from solved, and this should definitely remain a focus in future events. In the future, it would be desirable to have a data set that will allow a competition on detection of all 31 AUs, plus possibly a number of FACS action descriptors. Aside from addressing the detection of the activation of AUs, it would be a good thing to move toward the detection of the intensities and temporal segments of AUs, as it is these characteristics that prove to be crucial in many higher level behavior understanding problems.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments which were very helpful in improving the quality and presentation of this paper.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinene, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[2] N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis," Psychol. Bull., vol. 11, no. 2, pp. 256–274, Mar. 1992.

[3] A. B. Ashraf, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon, "The painful face—Pain expression recognition using active appearance models," Image Vis. Comput., vol. 27, no. 12, pp. 1788–1796, Nov. 2009.

[4] A. Asthana, J. Saragih, M. Wagner, and R. Goecke, "Evaulating aam fitting methods for facial expression recognition," in Proc. Int. Conf. Affective Comput. Intell. Interact., 2009, pp. 1–8.

[5] T. Baltrusaitis, D. McDuff, N. Banda, M. Mahmoud, R. El Kaliouby, P. Robinson, and R. Picard, "Real-time inference of mental states from facial expressions and upper body gestures," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 909–914.

Biography

Vutukuri Samba Siva Rao is a PG Student. He is doing Masters in Electronics and Communication Engineering from DRK Institute of Science&Technology, Hyderabad with specialization in VLSI&ES.

Dr.R.V.Krishnaiah received his B.Tech degree in ECE from Bapatla Engineering College. He received his M.Tech degree in Computer Science Engineering from JNTU and also M.Tech-(EIE) from NIT,Warangal. He did Ph.D (MIE, MIETE, MISTE) from JNTU Ananthapur.