A Survey on: Emotion Recognition with respect to Database and Various Recognition Techniques

Spurti Shinde
Pune Institute of Computer Technology

Sachin Pande
Assistant Professor
Pune institute of Computer Technology

ABSTRACT
Recognition and extracting various emotions and then validating those emotions from the facial expressions has become important for improving the overall human computer interaction. This paper reviews the literature on different aspects like different theories of emotions, methods for studying different images in the databases, different action units like outer brow raisers where the frontals and pars medal's facial muscles are studied. The paper reviews comparative techniques for automatically recognizing facial actions in sequences of images. The goal of this research is to show the comparison with Other AU Recognition Systems Comparison of selected facial expression recognition technique with different approaches on JAFFE database and Cohn-Kanade database. To study and evaluate their performance, using JAFEE and Cohn Kanade database. The basic five principal emotions to be recognized are: Angry, Happy, Sad, Disgust and Surprise along with neutral. Their recognition rate is obtained on all the facial expressions and observed comparatively.

Keywords
Face recognition technique, Action units, JAFEE Database, Emotions, Emotion detection, Cohn Kanade database.

1. INTRODUCTION
Emotion recognition is developing in the recent years because of different field like image processing and machine learning. 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 [4]. In this paper comparison of various recognition systems with respect to action units and various databases and comparative study emotion research is performed. The features considered while recognizing emotion can be static, dynamic, point-based geometric, or region based appearance.

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 and rapidly growing. In our opinion, the reviewed methods had a similar level of exposure and validation [2]. Our approach has been to present the existing methodologies to our readers in a way that will enable them to evaluate these approaches against their specific research objectives.

2. EMOTION RECOGNITION
Emotion recognition approaches can be divided into two groups based on the type of features used, either appearance based features or geometry-based features. Appearance features describe the texture of the face caused by expression, such as wrinkles and furrows and the different methods used in it are optical flow, Differential AAM and Texture and motion changes. Geometric features describe the shape of the face and its components such as the mouth or the eyebrows and the methods used in it are shape vectors, Facial animation parameter, distance and angular and trajectories. Furthermore, the detection of accurate human emotions is of vital importance for efficient human-computer interaction [15]. Numerous researches have explored this phenomenon which comprises 2D features, yet they are receptive to head pose, clutter, and variations in the lighting conditions. Human beings express various emotions in day to day interactions Understanding emotions and knowing how to react to people's expressions greatly enriches the interaction and will be helpful to know the intention behind the emotion

3. COMPARISON
3.1 AU Recognition Systems
Most automatic expression analysis systems attempt to recognize a small set of prototypic expressions, such as happiness, anger, surprise, and fear. Such prototypic expressions however occur rather infrequently. Human emotions and intentions are more often communicated by changes in one or a few discrete facial features [3]. During feature tracking, detailed parametric descriptions of the facial features are extracted. With these parameters as the inputs, a group of action units are recognized whether they occur alone or in combinations. The system has achieved average recognition rates of 96.4 percent (95.4 percent if neutral expressions are excluded) for upper face AUs and 96.7 percent (95.6 percent with neutral expressions excluded) for lower face AUs. The generalizability of the system has been tested by using independent image databases collected and FACS-coded by different research teams. Action Units (AUs) are the fundamental actions of individual muscles or groups of muscles. Action Descriptors (ADs) are unitary movements that may involve the actions of several muscle groups (e.g., a forward thrusting movement of the jaw). The muscular basis for these actions hasn't been specified and specific behaviors haven't been distinguished as precisely as for the AUs. Using Facial Action Coding System (FACS), human coders can manually code nearly any anatomically possible facial expression, deconstructing it into the specific Action Units (AU) and their temporal segments that produced the expression. As AUs are independent of any interpretation, they can be used for any higher order decision making process including recognition of basic emotions, or pre-programmed commands. FACS defines Action Units, which are a contraction or relaxation of one or more muscles. It also defines a number of Action Descriptors, which differ from Action Units in that the authors have not specified the muscular basis for the action.

Table 1. Comparison of action unit recognition system

Systems	Methods	Recognition rate	Treatment of AU combinations	Action units to be recognized	Databases
Current AFA System	Feature-based	88.5% (old faces)	No	AU 1,2,4, AU 5,6,7	Ekman-Hager
		89.4% (Novel Faces)	Yes /Yes		
		95.4%	Yes /Yes	AU 9,10,12,15 AU17,20,25 ,26, AU 27,23+24	Cohn-Kanade
		95.6%			
Barlett et al.[2]	Feature-based	85.3% (old faces)	No	AU 1,2,4, AU 5, 6, 7.	Ekman-Hager
		57% (Novel Faces)			
	Optic-flow	84.5%			
	Hybrid	90.9%			
Donator et al. [10]	ICA or Gabor wavelet	96.9%	No	AU 1,2,4, AU 5, 6, 7.	Ekman-Hager
	Others	70.3%-85.6%			
	ICA or Gabor	95.5%	Yes /No	AU 17,18,9+25	Ekman-Hager
	Others	70.3%-85.6%		AU 10+25,16+25	
Cohn et al.[8]	Feature - Tracking	89%	Yes /No	AU 1+2,1+4,4, AU 5,6,7	Cohn-Kanade (subset)
		82.3%	Yes /No	AU 12,6+12+25 AU 20+25,15+17 AU17+23+24,9+17 AU 25,26,27	
Lien Et al.[3]	Dense-flow	91%	Yes /No	AU 1+2,1+4, AU 4	Cohn Kanade (subset)
	Edge-detection	87.3%			
	Feature-tracking	89%	Yes /No	AU 1+2,1+4,4 AU 5, 6, 7.	
	Dense-flow	92.3%	Yes /No	AU 12,6+12+25 AU20+25,15+17 AU17+23+24,9+17	
	Feature-tracking	88%	Yes /No	AU 12,6+12+25 AU20+25,15+17 AU17+23+24,9+17 AU 25,26,27	
	Edge detection	80.5%	Yes /No	AU 9+17,12+25	

As shown in table 1 named" Comparison of action unit recognition system" the systems by considering different methods like feature based, optic flow, ICA or Gabor Wavelet on which various treatment of action unit combination is done .In table "No" means that no AU combination was recognized. "Yes/Yes" means that AU combinations were recognized and AUs in combination were recognizable individually. "Yes/No" means that AU combinations were recognized but each AU combination was treated as if it were a separate new AU [2].

Action units (AUs) represent the muscular activity that produces facial appearance changes defined in Facial Coding System by Ekman and Friesen [1].There are different methods used for various system implementation like feature based optic flow ICA or Gabor wavelets. Out of these by using ICA and Gabor wavelet we get maximum recognition rate up to 96.9% by considering each action unit separately like inner brow raiser, outer brow raiser, upper lid raiser, cheek raiser, lid tighter on Ekman Hager Database[10].

3.2 Comparison of selected facial expression recognition techniques

Table 2. Comparison of Selected Facial expression recognition system

Authors	Registration	Feature	Dynamic feature	Classifier
Jacobs and Davis	Not mentioned	Geometry: region based optical flow	yes	Rule based classifier
Essa and Pentland	3d model fitting	Geometry:3D motion and muscle models	yes	Euclidean norm
Wang et al.	Not mentioned	Geometry: B-Spline curve	yes	Euclidean norm

Hu et al.	Not mentioned	Geometry: variation of active shape model	yes	Probabilistic model
Valstar et al	Affine transform	Geometry: dynamics of 20 facial points	yes	Probabilistic actively learned support vector machine(SVM)
Pantic et al	Affine transform	Geometry:dynamics of 15 facial profile points	yes	Rule based
Donato et al	In-plane image transform	Appearance: Gabor wavelets	No	Nearest neighbor
Zhao and pietkainen	In-plane image transform	Appearance: BP on three orthogonal planes	Yes	SVM
Barlett et al	In-plane image transform	Appearance: Gabor wavelet	No	SVM and Adaboost
Wu et al.	In-plane image transform	Appearance: Gabor motion energy	Yes	SVM
Tian et al.	In-plane image transform	Hybrid: geometric + transient facial features	No	Neural network
Lucey et al	Active appearance Model(AAM)	Hybrid :2D shape + 2D appearance +3D shape	No	Nearest neighbor or SVM
Zhou et al.	AAM	Hybrid: geometry + SIFT	Yes	Multidimensional assignment algorithm

In machine learning, support vector machines [1] are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. There are various registration techniques in plane image and active appearance model. In table 2" Comparison of Selected Facial expression recognition system" it makes easy to work on Active Appearance Model and use hybrid feature using the nearest neighbor classifier or Support vector machine.

3.3 Comparison with different approaches on JAFFE database.

The JAFFE database contains 213 images with 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

Table 3. Accuracy on JAFFE using various features

Reference	Accuracy (%)	Features
Zhang et al. (1998)[3]	90.1	Geometry and Gabor
Bashyal and Venayagamoorthy (2008)[5]	90.2	Gabor and LVQ
Koutlas and Fotiadis (2008) [4]	92.3	Gabor filters
Liu and Wang (2006)[6]	92.5	Gabor filters
Oliveira et al. (2011)[7]	94.0	2DPCA with feature selection and SVM
Liao et al. (2006)[8]	94.5	LPB
Cheng et al. (2010) [9]	95.2	Gaussian process
Zhi and Ruan (2008)[10]	95.9	2D locality preserving projections

As one of the most important biometric techniques, face recognition has gained lots of attentions in pattern recognition and machine learning areas. On the same lines Emotion recognition has also gained popularity in the field of human computer interaction. Usually, a 2D facial image is represented as a feature point in the high dimensional feature space. It is perceptually structure can be characterized by using a small set of meaningful parameters. Thus, dimensionality reduction techniques are commonly used before recognition Table 3 shows the performance of different approaches reported in the literature on JAFFE database. Some of these results are not comparable directly as some authors exclude some classes of the problem. Accordingly as referring to table 3 using 2D Locality Preserving Projections (LPP) on the facial expression we can get accuracy up to 95.9%. 2D LPP is used for the dimensionality reduction, which detects intrinsic manifold structure of data and preserves local information.

3.4 Cohn-Kanade database

The Cohn-Kanade AU-Coded Facial Expression Database is for research in automatic facial image analysis and synthesis and for perceptual studies. Cohn-Kanade is available in two versions and a third is in preparation.

Version 1 (the original or initial release (Kanade, Cohn, & Tian, 2000)) includes 486 sequences from 97 posers. Each sequence begins with a neutral expression and proceeds to a peak expression. The peak expression for each sequence is fully FACS (Ekman, Friesen, & Hager, 2002; Ekman & Friesen, 1979) coded and given an emotion label. The emotion label refers to what expression was requested rather than what may actually have been performed. For validated emotion labels, please use version 2, CK+, as described.

Version 2, referred to as CK+, includes both posed and non-posed (spontaneous) expressions and additional types of metadata. For posed expressions, the number of sequences is increased from the initial release by 22% and the number of subjects by 27%. As with the initial release, the target expression for each sequence is fully FACS coded. In addition validated emotion labels have been added to the metadata. Thus, sequences may be analyzed for both action units and prototypic emotions. The non-posed expressions are from Amador, Cohn, & Reed. Additionally, CK+ provides protocols and baseline results for facial feature tracking and action unit and emotion recognition. Tracking results for shape and appearance are via the approach of Matthews & Baker. For action unit and expression recognition, a linear support vector machine (SVM) classifier with leave-one-out subject cross-validation was used. Both sets of results are included with the metadata.

3.4.1 Comparison with different approaches on Cohn-Kanade database.

Table 4 Accuracy on Cohn Kanade database using various features

Reference	Accuracy (%)	Features
Shan et al. (2005)[12]	79.1	LBP + template matching
Cohen et al. (2003)[11]	73.2	Gabor filter + SVM
Bartlett et al. (2003)[14]	86.9	Geometric + Tree-augmented-Naive Bayes
Bartlett et al. (2005)[14]	(Exp. II) 89.1	Gabor filter + SVM
Shan et al. (2009)[13]	88.9	LBP-based + SVM
Shan et al. (2009)[13]	86.8	Gabor filter + SVM

As shown in Table 4 "Accuracy on Cohn Kanade database using various features" shows the performance of different approaches reported in the literature on Cohn-Kanade database. It is observed that when second experiment was done using combination of Gabor filter and Support vector machine. The accuracy obtained was more i.e. 89.1% instead of geometric and Tree augmented Naive bayes in which it was 86.9%.

4. CONCLUSION

From the above discussion table 1 shows best recognition rate using ICA or Gabor wavelet. In which combination of action units were not considered and action units 1,2,4,5,6,7 were recognized from Ekman Hager .Table 2 shows that using active appearance model and by considering hybrid model and applying neural network classifier for preprocessing and feature extraction will be better approach. In table 3 we can observe that using 2D locality preserving projections we get accuracy upto 95.9% while by consider features like geometry and Gabor the accuracy which we get is 90.1% on Jaffe database .In table 4 we get the maximum accuracy upto 89.1% on Cohn Kanade Database, using the features like Gabor filter and SVM. This paper has made an attempt to put forth different techniques, databases used for the researchers in the area of Emotion Recognition. From this we can suggest in future using 2D Gabor filter and neural network we can try to increase the accuracy for recognizing the basic emotions like sad, happy, and angry.

5. REFERENCES

[1] Liao, S., Fan, W., Chung, C. S., & Yeung, D.-Y. (2006). Facial expression recognition using advanced local binary patterns. In International conference on image processing (ICIP) (pp. 665–668). IEEE.

[2] Liu, W., & Wang, Z. (2006). Facial expression recognition based on fusion of multiple Gabor features. In 18th International conference on pattern recognition 2006.

[3] Zhang, Z., Lyons, M. J., Schuster, M., & Akamatsu, S. (1998). Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. In Third IEEE international conference on automatic face and gesture recognition (pp. 454–459).

[4] Koutlas, A., & Fotiadis, D. I. (2008). An automatic region based methodology for facial expression recognition. In IEEE International Conference on Systems, Man and Cybernetics (pp. 662–666).

[5] Bashan, S., & Venayagamoorthy, G. K. (2008). Recognition of facial expressions using Gabor wavelets and learning vector quantization. Engineering Applications of Artificial Intelligence, 28, 1056–1064.

[6] Liu, W., & Wang, Z. (2006). Facial expression recognition based on fusion of multiple Gabor features. In 18th International conference on pattern recognition 2006 (pp.607 536–539).

[7] Oliveira, L. E. S., Koerich, A. L., Mansano, M., & Britto, A. S. Jr., (2011). 2d principal component analysis for face and facial-expression recognition. Computing in Science and Engineering, 13(3), 9–13

[8] Shih, F. Y., Chuang, C.-F., & Wang, P. S. P. (2008). Performance comparisons of facial expression recognition in JAFFE database. International Journal of Pattern Recognition and Artificial Intelligence, 22(3), 445–459.

[9] Liao, S., Fan, W., Chung, C. S., & Yeung, D.-Y. (2006). Facial expression recognition using advanced local binary patterns. In International conference on image processing (ICIP) (pp. 665–668). IEEE.

[10] Cheng, F., Yu, J., & Xiong, H. (2010). Facial expression recognition in jaffe dataset based on Gaussian process classification. IEEE Transactions on Neural Networks, 572 21(10), 1685–1690.

[11] Zhi, R., & Ruan, Q. (2008). Facial expression recognition based on two-dimensional discriminate locality preserving projections. Neurocomputing, 71, 1730–1734.

[12] Shan, C., Gong, S., & McOwan, P. (2005). Robust facial expression recognition using local binary patterns. In IEEE International Conference on Image Processing (pp. 370–373).

[13] Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing.

[14] Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing facial expression: Machine learning and application to spontaneous behavior. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 568–573).

[15] Alberto Battocchi, Fabio Pianesi, and Dina Goren- Bar. A first evaluation study of a database of kinetic facial expressions (dafex). In ICMI '05: Proceedings of the 7th international conference on Multimodal interfaces, New York, NY, USA, 2005. ACM.