Interpretation and Recognition of Dynamic Facial Action from the Image or Video

Pavan Ahire[1], Ranjana Dahake[2]

[1] ME Scholar, Computer Department, MET BKC IOE, Maharashtra, India
[2] Prof R.P. Dahake, Computer Department, MET BKC IOE, Maharashtra, India

ABSTRACT

Automated facial expression and recognition is a research subject on the basis of their interesting applications in the area of human and computer cooperation. The classification accuracy achieve by automatic system which uses static images as input have barrier by their image quality, lighting conditions and the direction of the detailed face. These problems can be overcome by fully automated facial expression detection and classification framework. a different classifier consist of particular sets of Random Forests that are equipped with support vector machine labellers. The system performs at actual-time rates under realize conditions, The dynamic acted still-image database is used for training purposes and expression-oriented video databases were used for testing. The quality evidence for emotional facial expression recognition conclude the main theoretical contribution to the field.

1. INTRODUCTION

Human face is elastic object that have various attributes related to the facial objects. Some behavior can extract human intention that exhibits some emotions. Tian et al specially pointed this out in their paper thorough 2011 review of the state of facial expression research [1]: "In the face expression literature, use of multiple viewpoint is rare; and relatively less attention has been focused on the problem of posture invariance". Over the past two years however, more work has been dedicated to this topic as low cost, high quality consumer cameras came bundled with computers and gaming consoles [2],[3].A totally automated complete facial expression detection and classification scheme is design and implemented in this paper. A typical facial expression recognition system Begins with the detection and acquisition of a face from an image or video. a commercial detector previously licensed by PittPatt [4].A set of random forests paired with support vector ma-chine labelers, it used a proprietary face detector and a novel classifier consisting With no intermediate human intervention. Various facial images and videos are used to recognize and detect expressions using expression labellers. The classification of facial expressions emoted by individuals in a spontaneous environment observed by video cameras. The major psychological premise is that externally exhibited facial expressions are an important window on internal emotions. First track and crop each face in each video frame, as well as to locate the center of the face. Yaw and roll angles are also determined. Once a face has been detected, normalize it with respect to brightness, contrast, size and pose. Subsequently by using feature analysis and dimensionality reduction to map the face into a space where expressions are identified using a classifier consisting of two stages, evaluation and labelling. The first involves a collection of binary random forests(RFs)[5] based on yaw bins, where each possible expression- yaw (or expression-intensity-yaw bin)combination is represented using an individual RF. The RF collection employed here is an entirely novel structure that benefits from the advantages of individual RFs, It breakdown a multi-label classification problem into a number of more manageable binary sub-problems. The second stage is an SVM labeller that takes the output of the RF collections and extracts a single, most likely, expression from all of the available possibilities.

1.1 RELATED WORK

De laTorre and Cohn have recently reviewed facial expression Research in their 2011 summary, focusing on state-of-the- art visual methods [6]. A major portion of this research in computer vision has been based on the contributions of Ekman and his various collaborators over the years [7].They have developed a model based on the premise that there exist six emotions, the communication of which can be "universally recognized "through the physical manifestation of facial expressions. These six emotions, originally selected by interaction with isolated tribes in Papua New Guinea [8] are happiness, surprise, fear, sadness, anger, and disgust at varying degrees of intensity. Ekman's "Big Six" [8] as the underlying emotion model. This gives us a set of dimensions (or axes), one for each emotion, and defines a so-called "emotion space". An emotion classifier using BU- 3DFE [9] is used as the training database. An expression classifier that uses a random forest evaluator combined with a support vector machine (SVM) labeller determines the emotion. The classifier permits the statistical collection of results, as opposed to just singular labels as is usually the case in the literature. Thus, near-continuous values or scores for each emotional dimension are achieved, with the granularity of the scale being controlled by the number of trees within each forest.

1.2 IMAGE PROCESSING

Alignment and orientation of images is of particular significance for appearance-based approaches where spatial coherence is of the utmost importance. Polynomial warping, described in [10] and [11], can provide this spatial coherence. Instead the coordinates of the facial landmarks detected by PittPatt to define a simple rectangular cropping window. The average dimensions of training samples sharing the same pose control the resizing of the detected faces. Thus, simple arithmetical operations requiring negligible processing time are involved in the alignment step.

1.3 SEQUENCIAL TRACKING

Classification of facial expressions in videos requires a method for combining information from frame to frame to produce a stable, reliable label. A form of temporal integration for this purpose in which a histogram of the resultant labels in a local temporal window surrounding each frame is analyzed and a single label chosen based on a popularity measure. The histogram of these single frame classification results is then used to produce the final single expression label.

1.4 INTER-EXPRESSION CORRELATION

Although many publications in the literature have adopted Ekman's theories, few have experimentally quantified the similarities between the six "universal" expressions. Extracting these correlation values identifies the difficulty of distinguishing each expression with respect to its counterparts.

1.5 INTRA-EXPRESSION CORRELATION

It providing new experimental evidence for the existence of strong relationships between intensities of the same expression, a topic not usually covered in the literature. In addition, evidence for the proximity of low intensities of different expressions is provided.

2. METHODOLOGY

The objective of this research is to develop a dynamic expression classifier capable of functioning at arbitrary yaw and roll and then evaluating its performance. However, they unable to find a source of unconstrained labelled video facial expression training data in the literature. Such a publicly available video database would need to contain a sufficient number of labelled faces at arbitrary poses. Therefore, the static BU-3DFE Database [9] is used for training, although, a video database would have been preferable.

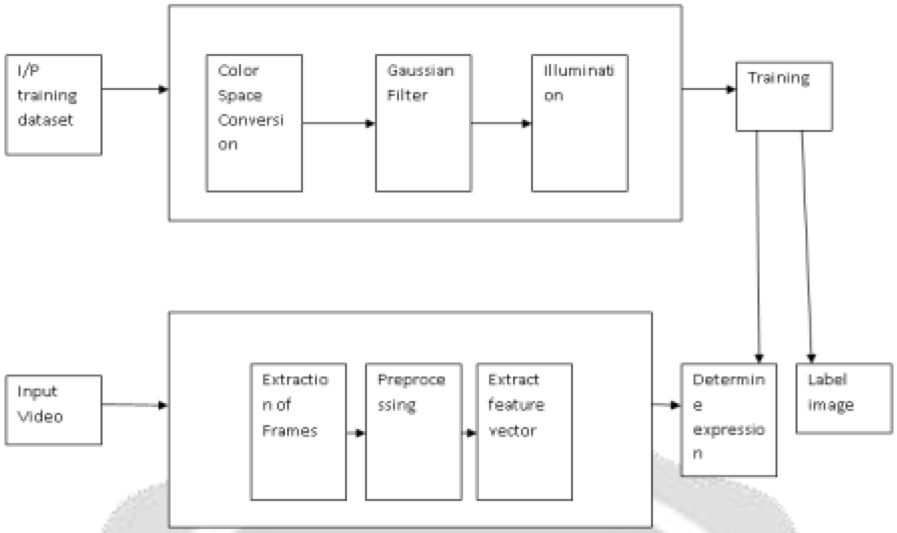

Figure -1: block diagram for automatic facial Expression system

There are various steps in which this system goes through are as follows:

1. Database preparation for training the classifiers.
2. Classifier training of feature vectors extracted from the training database.
3. Temporal integration of a video stream for accumulating and averaging classification results for individuals.
4. A testing strategy to evaluate the performance of the system.

2.1 Colour space conversion

A color space is a method by which we can specify, create and visualize color. A color is thus usually specified using three co-ordinates, or parameters. These parameters describe the position of the color within the color space being used. They do not tell us what the color is, that depends on what color space is being used.

2.2 Gaussian filter

Gaussian filtering is used to blur images and remove noise in detail. It also illustrate a probability distribution for noise or data. it is a smoothing operator that need to use the two dimensional Gaussian function. It is a low pass filter specially used for images

2.3 Illumination technique

It is specially used for preserving constant image quality Calculation of coherence or probability using side information with current cluster. After that Training of I/P dataset is start.Same thing is happen for input video from which frames are input from the videos for label image. This three sub steps are use for Determination of expressions in videos. This flow of Determination for label image is as follows: Extraction of frames Preprocessing Extract feature vector etc.

3. PERFORMANCE ANALYSIS

3.1 Still-Image Analysis

The contents of the publicly available image databases used for testing vary greatly in image size, image quality, camera characteristics, lighting, background, subject demographics, as well as the interpretation and presentation of the various expressions. Not all expressions were detected with equal accuracy. Happiness (98 percent) and surprise (90 percent) were the most identifiable expressions in still images, which are also confirmed by our experimental results. Also note that our experimental results for disgust were a departure from the general trend of human classification. During observation training of the classifier, that the Central yaw bins consistently experienced higher accuracies than other yaws. This was also obvious from the results for KDEF [13] and RaFD [14] that had used different databases for training and testing.

3.1 Video Analysis

The three video databases discussed in this section. Fluctuations in accuracy rates from database to database, as well as from feature to feature, were quite prevalent for the two video labeling schemes examined (schemes 2 and 3). This may indicate that the data are inseparable under either of the labeling schemes and more investigation into the nature of the poPDF vectors is necessary. The results illustrate the ability of different expression classifiers to identify certain expressions under the specific database's conditions. The results suggest that a practical application may require additional sensory inputs, such as, for instance, audio, to improve the outcomes in a real world situation. In addition, more emphasis could be placed on brightness and contrast Normalization at the image processing stage in order to Adapt to the existing environment.

4. CONCLUSION

This paper discusses a framework for classifying expressions in spontaneous situations. The classifier obtained was trained on the BU-3DFE image database and tested on other databases. The classification accuracies were found to be close in performance to the comparable platform found in [12]. Furthermore, our framework achieved real-time performance in a spontaneous environment, which has not previously been presented in the literature. Our biggest misgivings with respect to the theoretical foundation of this work were concerned with Ekman's "universality" of expressions. These qualms presented themselves at two distinct points in the research. The first point was the interpretation of expressions by posed database participants, and how these expressions related to their spontaneous, real-world counterparts. The second point deals with the classification of expressions in videos. Under a still image training strategy such as ours, successful video classification is highly dependent on the assumption that the progression from one "universal" expression (the source) to another (the sink) involves a sequence of intermediate expressions pertaining to the former or the latter. In reality, these intermediate expressions may contain elements of "non-universal" expressions, not captured by the Ekman model. A video classification system should intrinsically accommodate these intermediate expressions through training on progressions to and from apical 'universal' expressions. An even more integrated approach to classifying facial expressions probably requires multi-modal fusion through the amalgamation of sound, body posture and surrounding environmental data. These additional sensory sources build context, which is considered to be invaluable in the literature. We would expect that the use of sensory context would facilitate the ability to differentiate between expressions previously inseparable using a single medium such as a still image or video.

The conclusions above are a result of experiments with a new classification framework involving a combination of multiple RFs as classifiers followed by an SVM labeller. This expression classifier is new to the literature. The results presented here show that RF collections are the most significant aspect of the classifier and open the door to further research on this topic.

5. ACKNOWLEDGEMENT

I would like to express my special appreciation and thanks to my advisor prof R.P.Dahake, you have been a tremendous mentor to me.

6. REFERENCES

[1] .Y.-L. Tian, T. Kanada, and J. F. Cohn, “Facial expression recognition,”in Handbook of Face Recognition, 2nd ed., New York, NY, USA: Springer, 2011. pp. 487–519.

[2] .OMRON (2012). OMRON Develops real-time facial expression estimationtechnology[Online].Available:http://www.omron.com/media/press/2012/10/e1023.html.

[3] .B. Seddik, H. Maamatou, S. Gazzah, T. Chateau, and N. E. Ben Amara, “Unsupervised facial expressions recognition and avatar reconstruction from kinect,” in Proc. 10th Int. Multi-Conf. Syst., Signals Devices, 2013, pp. 1–6.

[4] .Pittsburgh, Pattern Recognition. (2011, Nov. 4) PittPatt SDK v5.2.2 Documentation[Online].Available: http://www.pittpatt.com/documentation.

[5] .L. Breiman, “Random forests,” Mach. Learn., vol. 45, pp. 5–32, 2001.

[6] .F. de la Torre and J. F. Cohn, “Facial expression analysis,” in Guide to Visual Analysis of Humans: Looking at People. New York, NY, USA: Springer, 2011.

[7] .P. Ekman, W. V. Friesen, and P. Ellsworth, Emotion in the Human Face: Guide-Lines for Research and an Integration of Findings. New York, NY, USA: Pergamon, 1972.

[8] .P. Ekman, “Universals and cultural differences in facial expressions of emotion,” in Proc. Nebraska Symp. Motivation, Lincoln, NE, USA, 1971, pp. 207–283.

[9] .Y. Lijun, W. Xiaozhou, S. Yi, W. Jun, and M. J. Rosato, “A 3D facialexpression database for facial behavior research,” in Proc. 7th Int. Conf. Autom. Face Gesture Recog., 2006, pp. 211–216.

[10].Z. Liu, Y. Shan, and Z. Zhang, “Expressive expression mapping with ratio images,” in Proc. 28th Annu. Conf. Comput. Graph. Interactive Techn., 2001, pp. 271–276.

[11].N. Arad, N. Dyn, D. Reisfeld, and Y. Yeshurun, “Image warping by radial basis functions: Application to facial expressions,” CVGIP: Graphical Models Image Process., vol. 56, pp. 161–172, 1994.

[12].W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, “Facial expression recognition using radial encoding of local Gabor features and classifier synthesis,” Pattern Recog., vol. 45, pp. 80–91, 2012.

[13].D. Lindquist, A. Flykt, A. €Ohman, “The karolinska directed emotional faces - KDEF” P. s. Dept. Clinical Neurosci., KarolinskaInstitutet, Ed., Solna, Sweden, 1998.

[14].O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, “Presentation and validation of the Radboud faces database,” Cognition Emotion, vol. 24, pp. 1377–1388, 2010.