Robust Facial Expression Recognition in the Presence of Rotation and Partial Occlusion

Diego Mushfieldt[1], Mehrdad Ghaziasgar[1], James Connan[2]
Department of Computer Science
University of the Western Cape[1], Private Bag X17, Bellville 7535
Tel: + (27) 21 959-3010, Fax: + (27) 21 959-3006
and Department of Computer Science
Rhodes University[2], P O Box 94 Grahamstown, 6140
Tel: + (27) 46 603-8291, Fax: + (27) 46 636-1915
email: {2721354, mghaziasgar}@uwc.ac.za[1]; j.connan@ru.ac.za[2]

***Abstract*—Facial expressions provide essential information about the emotional expressions of a signer in sign language recognition. The research presented in this paper investigates the effectiveness of facial expression recognition in non-frontal view facial images under the assumption that a signer's head may have slightly rotated. In South African Sign Language (SASL), partial occlusion of the face may occur when performing certain signs. Thus, this research aims to overcome the limitations of current systems which rely on controlled laboratory environments and discusses the development of a robust facial expression recognition system that can recognize facial expressions in non-frontal view and partially occluded facial images. At present, the face detection component of this research has successfully been implemented which is an essential step towards the accurate recognition of facial expressions.**

***Index Terms*—Facial Action Coding System, Facial Expression Recognition, Occlusion, Rotation, Support Vector Machine**

I. INTRODUCTION

Facial expressions carry essential information about the mental and emotional states of conversation partners [1]. Facial expression recognition is purely based on visual information obtained by facial movement and facial feature distortion. The visual information is categorized into abstract classes and classified [2].

All distinguishable facial movements can be measured by a comprehensive model known as, the Facial Action Coding System (FACS). FACS defines facial expressions in terms of the presence or absence of 44 unique muscle movements, or "action units" (AU's). [3]

The SASL research group at the University of the Western Cape is in the process of developing a machine translation system that can automatically translate between South African sign language and English [4]. Since some SASL signs contain facial expressions used to convey the emotional state of a signer [1], a robust facial expression recognition system is required.

Extensive research on facial expression recognition has been carried out by the SASL group. Whitehill [1] compared the recognition rates between two feature extraction methods — local and global segmentation. Sheikh [5] used various noise models — Gaussian noise, salt and pepper noise and speckle noise, to investigate facial expression recognition on noise degraded facial images. These empirical studies only consider frontal view and non-occluded facial images.

This research proposes a robust facial expression recognition system that can accurately recognize facial expressions in non-frontal view and partially occluded facial images.

The rest of the paper is organized as follows: Section II discusses related work; Section III describes the research goal; Section IV discusses the research methodology; the research is concluded in Section V.

II. RELATED WORK

The first subsection discusses the system developed by Hu *et al.* [6], which compares the recognition accuracies between non-frontal and frontal view facial expression analysis.

The next subsection discusses the system designed by Bourel *et al.* [7], which focuses on the recognition of facial expressions under the presence of partial occlusion.

A. *Hu et al.'s work*

The system uses a publically available 3D facial expression database created by Binghamton University, geometric salient facial points, and various classifiers to evaluate whether non-frontal view facial expression analysis can achieve higher recognition accuracy than existing frontal view facial expression methods.

The system uses a publically available 3D facial expression database containing six prototypical facial expressions consisting of 100 subjects — 60% female, 40% male of various ethnicity. Each subject performs seven expressions: neutral, happiness, disgust, fear, angry, surprise and sadness. These images are rotated at 5 different angles: 0, 30, 45, 60, and 90 degrees.

Geometric facial points are selected around the mouth, eyes and eyebrows at different rotation angles. The geometric 2D displacements of feature points between neutral and emotional expressions at corresponding angles are calculated.

The resultant displacement vectors are used as input to the following classifiers: Linear Bayes Normal Classifier, Quadratic Bayes Normal Classifier, Parzen Classifier, Support Vector Machine (SVM) with linear kernel, and K-Nearest Neighbor classifier.

A total of 1200 images are used consisting of 100 test subjects, each split into five groups of 20. Testing data consisted of the use of only one group, whereas training data consisted of the remaining four groups. Each test subject performed six emotions, which consisted of five intensity levels and rotation angles. The neutral expression was omitted. Experimentation showed that the SVM obtained the best recognition accuracy in comparison with other classifiers.

The system uses a 3D facial expression database and images are trained using different classifiers. Similar to this system, the proposed system will use an SVM for facial expression recognition since results show that the SVM obtained the highest recognition accuracy. However, feature points in this system are manually selected and not tracked, whereas the proposed system will automatically select feature points and track those points. The proposed system will also test rotation angles in a similar manner to this system.

B. Bourel et al.'s work

The system uses recognition techniques such as localized representation of facial features and data fusion, to overcome the limitations of existing occlusion-invariant techniques, which rely on laboratory controlled environments.

A total of 12 facial feature points are manually selected at the eyebrows, nostrils and mouth regions respectively.

The tracker, which is based on the Kanade-Lucas tracker, is able to track and recover any of these feature points lost due to illumination variations and head orientations.

The points are geometrically related to each other. Motion information relative to the start and endpoints of the dynamic part of the sequence are recorded by these vectors.

When various sources of information are independent, data fusion improves the accuracy of pattern classification. The recognition score for each known expression is obtained by each local expert classifier. The classifiers are combined to produce a final classification result.

100 video sequences were used for testing, in sets of 25 sequences per expression class. Four expressions were tested: anger, joy, sadness and surprise. Testing and training sets were partitioned according to the leave-one-out methodology. Testing data consisted of 30 subjects from the Cohn-Kanade AU-coded facial expression database at Carnegie Mellon University. Missing facial regions simulated occlusion regions of the: mouth, upper face, or left/right half of the face. Experimentation reveals that occlusion of both left and right sides of the face achieved the same results. The 'anger' and 'surprise' expressions are robust to occlusion of the mouth. Results show that 'sadness' is not robust to mouth occlusion. All expressions are robust to occlusions of the other parts of the face with an average accuracy of 80%.

The system overcomes occlusion using a combination of local expert classifiers. In contrast to [6], K-Nearest Neighbor classifiers are used instead of SVMs. However, research carried out by Hu *et al.* suggests that the use of SVMs will certainly increase the recognition accuracy. Similar to this system, the proposed system will test occlusion in the same manner, where missing facial regions simulate partial occlusion.

III. RESEARCH GOAL

The aim of this research is to develop a robust facial expression recognition system that can accurately recognize facial expressions in non-frontal view, as well as partially occluded facial images.

IV. RESEARCH METHODOLOGY

A positivist philosophical approach will be used and the software design will follow the iterative model. A framework for facial expression recognition can be broadly categorized into three components: face detection, facial feature extraction and expression recognition.

1. In order to accurately classify facial expressions, the location of a signer's face must be determined. The face detection method used in this research is based on skin detection and was developed by Achmed *et al.* [8]. At present, this component of the facial expression recognition framework has successfully been implemented and detects faces at any orientation.
2. In order to track facial feature points, the tracker in [7] will be used.
3. Finally, an SVM will be used to recognize facial expressions instead of the K-Nearest Neighbor classifier since research in [6] shows that an SVM performs much better than other classifiers used for facial expression recognition.

V. CONCLUSION

This research proposes a robust facial expression recognition system that is invariant to partial occlusion and non-frontal view facial images. The motivation for such a robust system was clearly stated. This implementation can improve the recognition of facial expressions in real-world scenarios, specifically for the use of sign language translation.

REFERENCES

[1] J. Whitehill, "Automatic Real-Time facial expression recognition for signed language translation," University of the Western Cape, South Africa, MSc Thesis 2006.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[3] P. Ekman and E. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. New York, United States of America: Oxford University Press, 1997.

[4] M. Ghaziasgar and J. Connan, "Investigating the Intelligibility of Synthetic Sign Language Visualization Methods on Mobile Phones," in *Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, South Africa, 2010, pp. 86-92.

[5] M. Sheikh, "Robust Recognition of Facial Expressions on Noise Degraded Facial Images," University of the Western Cape, South Africa, MSc Thesis 2011.

[6] Y. Hu et al., "A Study of Non-frontal-view Facial Expressions Recognition," in *Pattern Recognition*, 2008, pp. 1-4.

[7] F. Bourel, C. Chibelushi, and A. Low, "Recognition of facial expressions in the presence of occlusion," in *Twelfth British Machine Vision Conference*, 2001, pp. 213-222.

[8] I. Achmed and J. Connan, "Upper Body Pose Estimation towards the translation of South African Sign Language," University of the Western Cape, Cape Town, 2010.

Diego Mushfieldt is currently an M.Sc student at the University of the Western Cape. He is currently doing research on sign language synthesis and novel communication applications on mobile interfaces for the Deaf and hearing impaired.

Mehrdad Ghaziasgar is the project manager of the South African Sign Language (SASL) research group. His interests are internet programming, mobile computing, computer vision and machine learning.

James Connan heads up the South African Sign Language (SASL) research group. His interests are computer vision and machine learning.