Study on Aging Effect on Facial Expression Recognition

Nora Algaraawi, Tim Morris

Abstract— **Automatic facial expression recognition (AFER) is an active research area in computer vision. However, aging causes significant alterations in the facial appearance of an individual (as do health and lifestyle). Changes in facial structure due to aging can affect the appearance of facial features resulting in poorer ability of humans and machines to recognize an individual's expression. This has not been studied in depth yet. In this paper, we describe how the changes in facial features across age affects an AFER system's performance using shape and texture information. We propose a framework to analyze and recognize various expressions from three kinds of feature space: shape features, texture features and appearance features. We extract features in two ways: from the whole face and from patches surrounding the eyes and mouth. The extracted features are trained using a kernel support vector machine (SVM) to recognize the expression. The FACES database from the psychology society is introduced to the computer vision community and used for our study. Experimental results demonstrate that aging has a big effect on the shape, texture and appearance features of face and expression. By training age appropriate classifiers the accuracies across groups and within group are about 99.25% and 100% respectively.**

***Index Terms*— aging effect on facial expression recognition, Facial expression recognition, facial expression analysis, facial expression recognition across the life span.**

I. INTRODUCTION

Facial expression recognition has become an active research area in recent years, since facial expression is one of the most powerful ways that people coordinate conversation and communicate emotions and other mental, social and psychological cues [2, 7].

Moreover, technologies that can interpret and respond to facial expressions automatically could be used in a wide range of applications. For example, in pharmacology, drug trials could be conducted more efficiently by analyzing expressions than by completing questionnaires [3]. Online teaching systems might adapt according to the students' expressions in the way good teachers do [4]. They could be used to assess the lassitude of drivers and pilots, which might be used to decrease the number of accidents.

Manuscript received March 18, 2016; revised March 31, 2016. The author is currently funded by The Higher committee for Education Development in Iraq (HCED).

Nora Al-Garaawi is a PhD Student at the School of Computer Science, The University of Manchester, Manchester, M13 9PL, UK.; (nora.al-araawi@postgrad.manchester.ac.uk).

Tim Morris is a Lecturer with the School of Computer Science, The University of Manchester, Manchester, M13 9PL UK (e-mail: tim.morris@manchester.ac.uk).

Facial expression is effected by many factors such as age, gender and type of expression [5]. Emotion perception has been found to vary with age-congruence between sender and receiver. Further, affective expressions of older individuals appeared harder to decode, owing to age-related structural changes in the face which supports the notion that the wrinkles and folds in older faces actually resemble emotions. Therefore older subjects may have poorer ability to recognize an individual's emotion at different ages. Fig. 1 shows the changes in the facial structure across ages.

To our knowledge, only one study has considered the effect of age on AFER [8], others have investigated subjects with a restricted age range. A key challenge we face is that the appearance of different expressions overlaps between the different age groups.

This paper presents our preliminary results regarding facial expression recognition taking into account the age of the subject. We proposed to extract and use shape, texture, and appearance features using well-known approaches: Active Shape Model (ASM) [9], Active Appearance Model (AAM) [10, 11] and Local Binary Pattern (LBP) [12]. Features are classified using the kernel Support Vector Machine (SVM) [13].

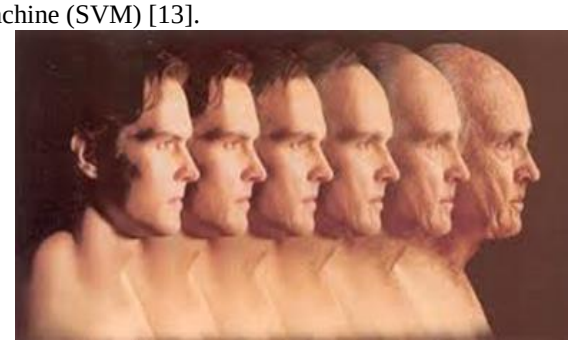

Fig. 1. The changes in facial structure with age [5].

Fig. 2. Some examples from the FACES database [1].

The reminder of this paper is organized as follows. In section 2 we describe the database used in this paper. Section 3 discusses the methods and the framework. Section 4 reports our experimental results and section 5 presents conclusions and future work.

II. DATABASE

The FACES database is used in this work. This database involves 171 people showing six expressions (anger, disgust, fear, happy, neutral and sad). They are divided into three groups according to their age (young: 19 to 31 years old, middle-aged: 39-55, and older: 69-80). Each individual shows two examples for each expression. In total the database consists of 2052 frontal face images [1]. Some examples from the database are shown in Fig. 2. See table 1 for more detail about the FACES databases.

Table 1: Analysis of FACES database

Gender	Age Group (Age range)			
	Young (19-31)	Middle (39-55)	Old (69-80)	Total (19-80)
Male	29	27	29	85
Female	29	29	28	86
Total	58	56	57	171

III. METHODS

Expressions are generated by muscles' movements. These will not only lead to changes in shape but also to changes in the visual appearance of the face. It should therefore be possible to classify facial expression by analyzing the shape and/or texture properties. We propose to study how aging will affect the face (shape and texture) and how that will affect the expression recognition.

We used the ASM to assess how the shape is affected by aging. Then, we have taken two approaches to assessing the facial appearance: we have used the AAM that captures the shape and appearance of objects, and compared it to a method that incorporates a face feature detector and LBP. Fig. 3 shows the main elements of our approach and indicates information flow within the system.

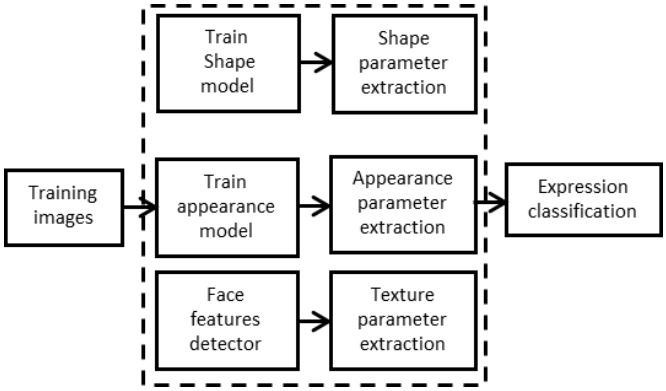

Fig. 3. The framework of the proposed AFER system.

A. Active Shape Model

We propose firstly to model the shape variations of faces using ASM. It captures the variability of shape of the object of interest based on a set of pre-annotated training images, where key landmark points are labeled. Then, any valid shape can be represented as a combination of the significant modes of variation [9]. In this paper, the modes' weights will be used to classify the facial expression.

The faces were labelled using two schemes. The first used 76 points following Cootes' original description, the left image of Fig 4. We found that the face shape was hard to describe accurately in the older groups, so an alternative scheme using 61 points was also used (the right image of Fig 4).

An age specific shape model (SM) and a general SM were created using both labeling schemes. The shape parameters were input to kernel SVM using ten-fold cross validation to assess the models' accuracies.

B. Active Appearance Model

The second method we propose to use is the active appearance. This combines the shape parameters with grey-level distributions that surround the landmarks. The appearance model (AM) provides a compact statistical representation of both the shape and appearance of the face image, although the appearance is described only in the vicinity of the landmarks [10, 11]. Again, the models' parameters can be used to recognize an expression.

The same two labeling schemes and the same strategy for estimating the models' accuracies were employed.

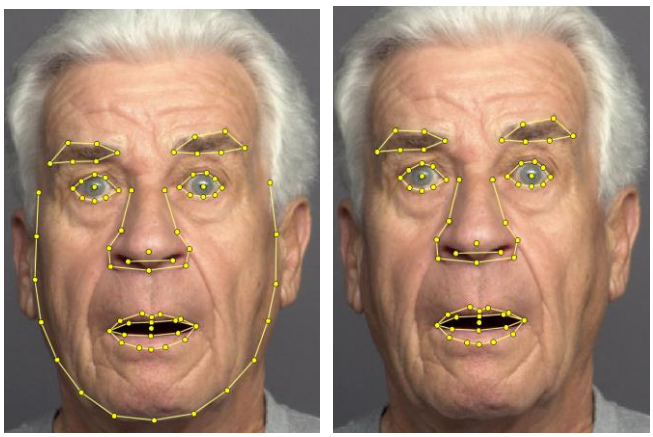

Fig. 4. Example face images annotated with 76 or 61 landmarks.

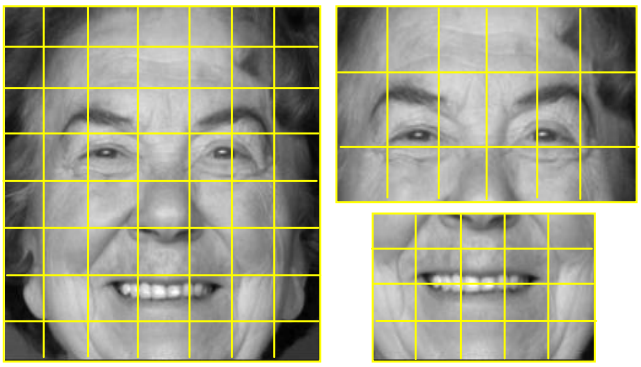

Fig. 5. Face, eyes and mouth regions and their subdivision.

C. Local Binary Pattern (LBP)

The final method we propose is LBP, a common method of quantifying texture [12]. It has two properties that make it attractive for characterizing texture: its invariance to any

monotonic grey level change e.g. those caused by illumination change, and its computational simplicity.

In its simplest form, a pixel's LBP is computed by systematically comparing its value to its eight neighbours' values. A 1 is recorded if the neighbour is greater than the pixel. The eight values constitute a value in [0, 255]. A region is characterized by the probability distribution of the LBP values see [12] for more detail.

Our feature extraction method with LBP was by:

1) The face was detected, then the eyes and mouth within the face area using Viola Jones detectors [14].
2) The face, mouth and eyes regions were divided into non-overlapping areas (Fig. 5) and an LBP distribution were calculated. A sub-histogram is computed for each smaller region and concatenated into one histogram.

Two schemes for representing the face were investigated, one using the whole face data, the other by concatenating the eyes and face region distributions which greatly contribute to facial expression. Again, age specific and a general model were created. Features were assigned to expression classes using kernel SVM and evaluated with 10 fold cross-validation.

IV. EXPERIMENTAL RESULTS

Our experiments are concerned with answering two questions: is expression recognition more effective if an age specific model is used? Can accuracy be improved by excluding portions of the face that may not contribute to expressions?

A. Active Shape Model Results

Shape models were built to represent each age group and the whole data set using 76 and 61 landmarks: eight models in total, 4 models using 76 points and 4 model using 61 points (one for each age group and one from all the data). Fig. 6, 7 and 8 show examples of the 76 point ASM Young model being fitted to Young, Middle and old data, they demonstrate how the model successfully fits to the correct data and how it misses the incorrect data. The reason is that by using ASM instances of models can only deform in ways found in the training set.

We further assessed the accuracy of the models in recognizing expressions. The extracted shape models parameters were classified using ten-fold cross validation and a kernel SVM. Results are presented in Table 2 for the 76-point models and Table 3 for the 61-point models. It will be seen that the age specific model is most appropriate for classifying features, although the global model does capture all the variation in the data. The 61-point models are probably no better than the other, but they are quicker to compute.

In conclusion, we can state that aging has an influence on the face's shape and affects the accuracy of expression recognition.

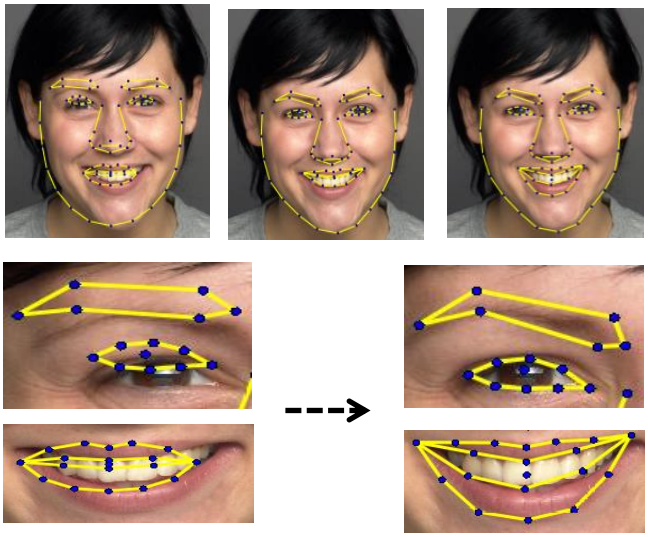

Fig. 6. Evaluation of Young ASM to Young data. Top row: after 1, 2 and final iterations of alignment. Bottom row: the alignment of the eye and mouth regions.

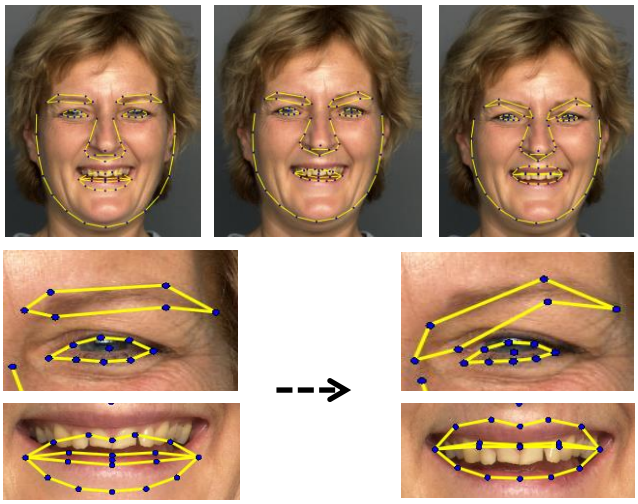

Fig. 7. Evaluation of Young ASM to Middle data. Top row: after 1, 2 and final iterations of alignment. Bottom row: the alignment of the eye and mouth regions.

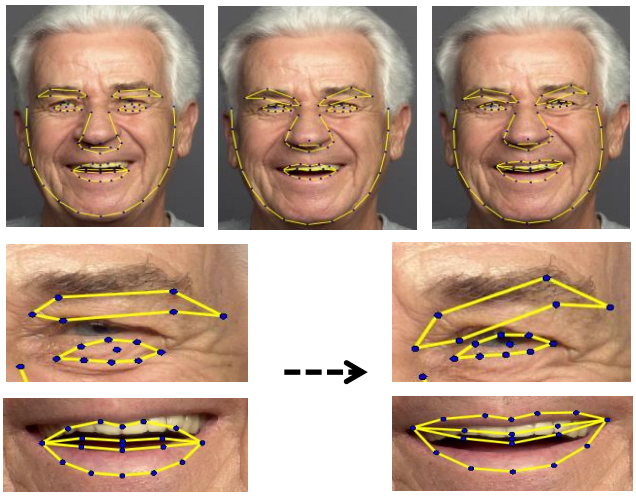

Fig. 8. Evaluation of Young ASM to old data. Top row: after 1, 2 and final iterations of alignment. Bottom row: the alignment of the eye and mouth regions.

Table 1: Classification Accuracies, 76-Point SM

Training Group	Testing Group			
	Young	Middle	Old	All Groups
Young	94.06	91.96	86.54	90.83
Middle	92.81	94.91	87.86	91.91
Old	90.94	93.00	90.58	91.42
All Groups	95.40	94.79	91.08	93.55

Table 2: Classification Accuracies, 61-Point SM

Training Group	Testing Group			
	Young	Middle	Old	All Groups
Young	94.76	92.70	87.57	91.71
Middle	92.81	95.26	87.13	91.81
Old	91.52	93.15	90.29	91.86
All Groups	95.54	95.23	91.52	93.68

B. Active Appearance Model Results

The appearance model (AM) adds a texture measurement to the SM. Fig. 9 illustrates the first three modes of variation of the 61-point model.

As in section A, eight appearance models were built and used to classify the expressions using the same methodology as before. Tables 3 and 4 present the recognition results.

Again we see good accuracy where a model is trained and tested on subjects from the same age groups. There is also little difference in accuracy between the two models, but the 61-point models are obviously smaller, suggesting that the data that is not represented does not provide useful information.

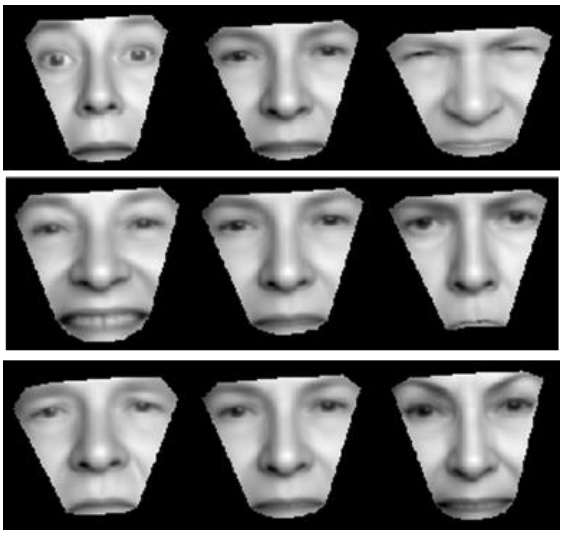

Fig. 9. The first three modes of the 61-point AM model.

In conclusion, we can state that aging has an influence on the face's appearance and affects the accuracy of expression recognition.

C. Local Binary Pattern Results

The hypothesis we propose to test is that the texture of regions around the eyes and mouth will provide as good discrimination between expressions as does the whole face.

Table 3: Classification Accuracies, 76-Point AM

Training Group	Testing Group			
	Young	Middle	Old	All Groups
Young	100.00	94.34	87.86	94.10
Middle	96.26	99.71	91.81	95.95
Old	94.10	94.79	98.98	95.85
All Groups	99.42	99.25	97.66	99.02

Table 4: Classification Accuracies, 61-Point AM

Training Group	Testing Group			
	Young	Middle	Old	All Groups
Young	100.00	93.15	86.54	93.27
Middle	96.55	99.82	89.47	95.32
Old	96.69	99.40	97.69	96.29
All Groups	99.51	99.40	97.69	99.25

As described above, we used the Viola-Jones detectors [13] to locate the face, then the eyes and mouth in each image. The LBP descriptor for the face, and the eyes and mouth were computed. An image of LBP values is given in Fig. 10. It is apparent that the regions that have been selected have extended beyond the face and into the background or hairline. This is an unfortunate consequence of the necessity of having a rectangular region of interest.

Again, kernel SVM classifiers were trained using the young, middle, old and the whole data set, for the whole face and the eye/mouth regions: eight classifiers in all. Their accuracy was again assessed by ten-fold cross validation. Results are presented in tables 5 and 6.

It is apparent that this technique is less able to generalize than the Appearance Models. It is also obvious that the specific age group models perform significantly worse on other age groups. However, the specific models perform very well on their targeted age group. There is a marginal improvement by using the face patches rather than the whole face, this improvement might be partially masked by the patches including areas that are not parts of the face.

In conclusion, we can state that aging has an influence on the face's texture and affects the accuracy of expression recognition

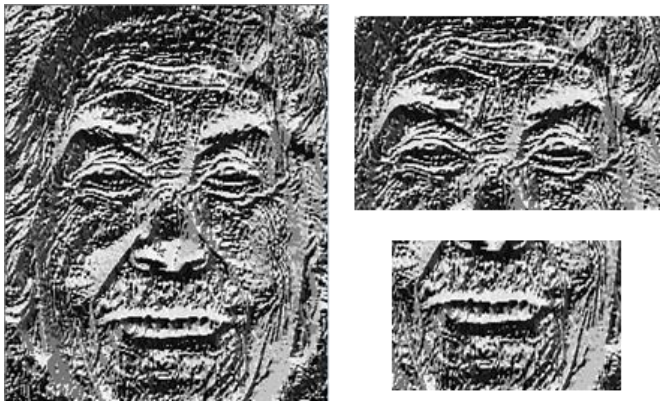

Fig. 10. LBP descriptors of figure 5.

Table 5: Classification Accuracies, LBP Whole Face

Training Group	Testing Group			
	Young	Middle	Old	All Groups
Young	100.00	83.92	59.79	78.55
Middle	92.24	100.00	81.28	87.37
Old	77.01	83.03	99.98	81.28
All Groups	100.00	100.00	94.44	95.21

Table 6: Classification Accuracies, LBP Face Patches

Training Group	Testing Group			
	Young	Middle	Old	All Groups
Young	100.00	81.54	57.16	78.31
Middle	86.92	100.00	71.78	83.38
Old	67.67	79.76	100.00	76.90
All Groups	100.00	100.00	94.20	95.21

D. Summary

Fig. 11 presents the within-group accuracies. From this we conclude that the pure texture measurements are able to classify expressions accurately, provided we use an age-specific model. This would suggest that the face shape measurements made in the SM and AM do not completely describe the data.

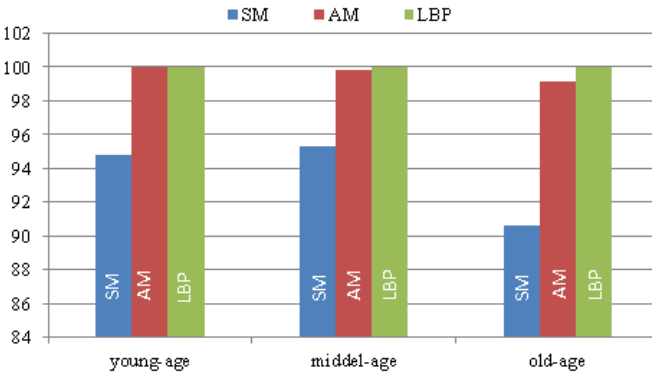

Fig. 11. Within Group Classification Accuracies

Fig. 12 presents the accuracies obtained using a model created from the whole data set. We conclude that the AM is best able to generalize across this data.

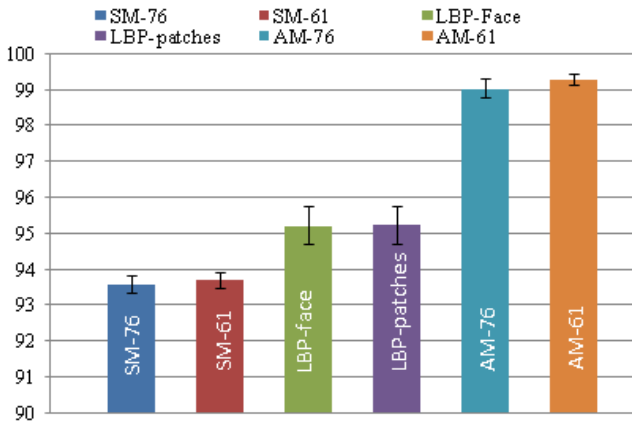

Fig. 12. Cross-Group Accuracies

We have also compared our methods against previous work [8] and human observers [1]. Table 7 summarizes the results.

Table 7: Accuracies of expression recognition by humans, reported by Guo et al [8] and our methods.

Method / group	Within Group	Cross Group
Human-perception		80.8
Guo et al. [8]	97.85	97.89
proposed method	100	99.25

E. Aging Effect Analysis

The results from all the experiments demonstrate that aging has significant effects on the appearance of expressions. In this section we attempt to explain and analyze the reasons for this.

[15] Claim that the main reason that expressions alter as we age is not only the presence of wrinkles but also the reduction in the elasticity of the facial muscles which leads to a change in facial appearance. For example, the fold between the cheek and upper lip can be appear in the happy expression of young people and the neutral expression of old people. Fig. 13 shows the responses of the LBP, it goes some way to explaining the poor performance of cross group classifiers.

In addition, having examined all the images in the FACES database we found that most of the young people show more exaggerated expressions than do the older people. Further, there are extraneous factors that affect a person's apparent age and further confuse the classifiers. This is consistent with psychology studies [16] [17]. As an example, see Fig. 14.

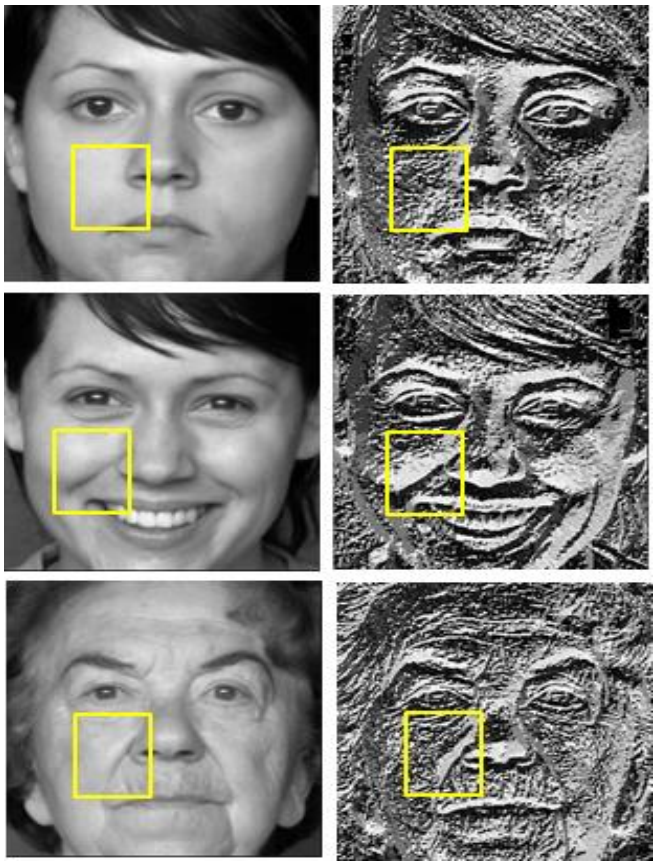

Fig. 13. A comparison of LBP response for expression and aging of the young-neutral, young-happy and old-neutral samples

Fig. 14. Top row: neutral expressions, bottom row: disgust. Left column: a young example, right column and old example.

V. CONCLUSIONS AND FUTURE WORK

In summary, the findings from this study extend the research regarding facial expression recognition by studying the effect of ageing on facial expression recognition. Our experiments highlight that ageing effects should be taken account of when trying to recognize expressions. By taking this into account the solution to expression recognition will be more general and more effective.

Model based approaches such as ASM, AAM and LBP have been proposed to analyze the effect of aging on the shape, texture and appearance of facial features and in consequence on the performance of expression recognition. The features derived by these methods were extracted from the whole face and from patches of the face that we believe are more responsible for the expressions. Our results demonstrated that the shape, texture and appearance of facial features change across the life span gradually. Therefore, the shape, texture and appearance and the effects of aging on them should be taken into account when trying to recognize expression.

A multiclass support vector machine has been used in our study to compare the extracted features. The results show that the highest accuracy when taking aging effects into account have been achieved using the appearance features from AM, about 99.25%. LBP has successfully recognized the expression in specific age groups with 100% accuracy. Using data from specific regions of the face does not degrade the recognition rates, supporting our hypothesis that the excluded regions do not contribute much, or at all, to the expressions.

We intend to evaluate and quantify the results more systematically using manifold learning approaches in the future. We also intend to investigate methods of automatically estimating age, in a combined age and expression recognition system.

References

[1] Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods*,*42*(1), 351-362.

[2] Ekman, P., & Friesen, W. V, 1987. *Facial Action Coding System*, Palo Alto, California: University of California Medical Centre, San Francisco.

[3] Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., ... & La Torre, F. D. (2009, September). Detecting depression from facial actions and vocal prosody. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (pp. 1-7). IEEE.

[4] Bartlett, M., Littlewort, G., Wu, T., & Movellan, J. (2008, September). Computer expression recognition toolbox. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on* (pp. 1-2). IEEE.

[5] Esme, B. and Sankur, B., 2010. Effects of aging over facial feature analysis and face recognition. *Bogazici Un Electronics Eng Dept*.

[6] Malatesta CZ, Fiore MJ, Messina JJ, 1987. Affect, personality, and facial expressive characteristics of older people. Psychology and Aging, 2(1), 64–69.

[7] Hareli, S. and Hess, U., 2012. The social signal value of emotions. *Cognition & Emotion*, *26*(3), pp.385-389.

[8] Guo, G., Guo, R., & Li, X. (2013). Facial expression recognition influenced by human aging. *Affective Computing, IEEE Transactions on*, *4*(3), 291-298.

[9] Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, *61*(1), 38-59.

[10] Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), 681-685.

[11] Cootes, T.F. and Kittipanya-ngam, P., 2002, September. Comparing Variations on the Active Appearance Model Algorithm. In *BMVC* (pp. 1-10).

[12] Ojala, T., Pietikäinen, M. and Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(7), pp.971-987.

[13] Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory* (Vol. 1). New York: Wiley.

[14] Viola, P., Way, O.M. & Jones, M.J., 2004. Robust Real-Time Face Detection. *international Journal of computer vision*, 57(2), pp.137–154.

[15] "Aging of the Face," http://www.face-and-emotion.com/dataface/facets/aging.jsp, 2013.

[16] Ebner, N.C. and Johnson, M.K., 2009. Young and older emotional faces: are there age group differences in expression identification and memory?.*Emotion*, *9*(3), p.329.

[17] N. Ebner and M. Johnson, 2010. "Age-Group Differences in Interference from Young and Older Emotional Faces," *Cognition and Emotion*,. 24(7), pp. 1095-1116.