Selecting, Optimizing and Fusing 'Salient' Gabor Features for Facial Expression Recognition

Selecting, Optimizing and Fusing 'Salient' Gabor Features for Facial Expression Recognition

Ligang Zhang, Dian Tjondronegoro

Faculty of Science and Technology, Queensland University of Technology,
2 George Street, Brisbane, 4000, Australia
ligzhang@gmail.com, dian@qut.edu.au

Abstract. This paper describes a novel framework for facial expression recognition from still images by selecting, optimizing and fusing 'salient' Gabor feature layers to recognize six universal facial expressions using the K nearest neighbor classifier. The recognition comparisons with all layer approach using JAFFE and Cohn-Kanade (CK) databases confirm that using 'salient' Gabor feature layers with optimized sizes can achieve better recognition performance and dramatically reduce computational time. Moreover, comparisons with the state of the art performances demonstrate the effectiveness of our approach.

Keywords: Facial expression recognition, Gabor filter, $(2D)^2PCA$, KNN.

1 Introduction

Facial expression recognition (FER) is an active area and has been increasingly given much attention in recent years due to its potential to be applied into a wide range of areas, including human-computer interaction, video surveillance, video indexing and summarization. To date, a robust FER is still a challenging issue due to facial image variations, such as illumination, rotation and occlusion.

FER method can be classified into four categories: motion-based, feature-based, model-based and appearance-based approaches. Appearance-based is the most effective approach to handle facial image in real situations since it is insensitive to in-plane rotation and illumination variations, particularly Gabor filter. However, there are three weaknesses in the use of Gabor filter which need to be overcome, including redundant information within the neighboring frequencies [1]; expensive computation [2]; and different channels have different contributions on recognition performance [3]. In this paper, we will address these problems by selecting 'salient' Gabor filters. There have only been few studies on the 'salient' Gabor features selection, which can be categorized into three groups: 1) Point based approach [4, 5] which extracts Gabor features based on fiducial points of a face grid. However, its recognition performance is dependent on the accuracy of the automatically selected and located fiducial points, which is still a challenging task. 2) Feature based approach which performs Gabor filters on facial images and selects the 'salient' features using feature selection algorithms such as Adaboost [6, 7], genetic programming (GP) [8] and zero norm [9].

The original publication is available at:
http://www.springerlink.com/content/k2252wwk2t1343h7/fulltext.pdf

Although it overcomes the drawback of point based approach, it still requires accurate face location. 3) Channel based approach [3] which aims to select a subset of Gabor channels corresponding to different scales and orientations. Unlike the other two methods, this approach eliminates the requirement of point location at the cost of losing expressional information in unselected channels. The selection can be specifically optimized for each expression [10] or overall performance [1].

In this paper, we propose a channel based approach that selects, optimizes and fuses a set of 'salient' Gabor filters for effective FER from still images. We extend Gabor filters from 5 to 18 scales and adopt $(2D)^2$PCA instead of PCA for dimension reduction. The selection of feature layers and the determination of their optimized sizes are automatically processed based on the recognition performance of an image set. The selected 'salient' layers are fused for six universal expressions recognition, including anger AN, disgust DI, fear FE, happy HA, sadness SA and surprise SU, using the K nearest neighbor (KNN) classifier.

The main contributions are as follows: 1) We propose a novel and automatic approach to select and optimize 'salient' Gabor features. To the best of our knowledge, our approach is the first attempt to exploit the selection of 'salient' Gabor features from the aspect of scale, orientation and size. Meanwhile, our approach also is the first one to explore the way of determining the optimized sizes of feature matrixes. 2) We investigate the recognition performances of KNN using K values ranged from 1 to 14. Our results indicate that the best performance is obtained when K equals to 1. 3) We use $(2D)^2$PCA for dimension reduction. Our results show that it only takes a small proportion of the overall computational time. 4) We confirm that using 'salient' features can lead to a better performance with dramatically less computational time than using all features. 5) We present results to confirm Littlewort's finding [7] that useful emotional features are distributed in a wide range of Gabor feature scales. 6) We use a comprehensive evaluation to demonstrate that "sad" contributes to most of the misrecognitions, while "surprise" is the easiest facial expression to be correctly recognized for both JAFFE and CK databases.

The rest of the paper is organized as follows. Section 2 describes in details the proposed framework and each step. Section 3 shows the performance evaluations using three types of comparisons, namely approach using all features, computational time and state of the art performances. Finally, conclusions are drawn in section 4.

2 System Framework

The proposed framework as shown in Fig. 1 is composed of five steps: pre-processing, Gabor features, $(2D)^2$PCA, layer selection and layer fusion. During pre-processing, face images are cropped and scaled into a resolution of 110*110 pixels. These images are then passed through 9 bands, 2 scales, and 4 orientations Gabor filters. In this paper, we define a *layer* as a Gabor feature representation with different bands, scales and orientations. These layers are processed by $(2D)^2$PCA for dimension reduction, which produces feature matrix layers with the same bands, scales and orientations, but smaller sizes. Layer selection is then automatically achieved based on the performance of an image set to choose the most 'salient' feature matrix layers and

decide their optimized sizes. Finally, the 'salient' optimized layers are fused for recognizing the six universal expressions using the KNN classifier.

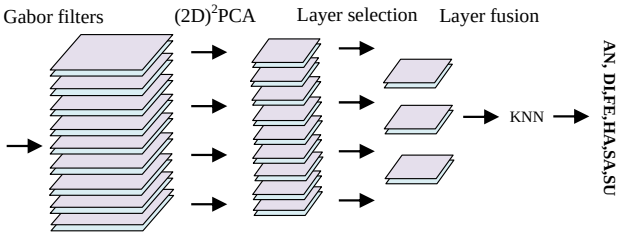

Fig. 1. Flow chart of the proposed framework.

2.1 Gabor Features

Gabor filters have been successfully applied to a wide range of fields, such as face recognition [11] and fingerprint identification [12]. In this paper, 2D Gabor filter is adopted and it can be mathematically expressed as:

$$F(x,y) = \exp\left(-\frac{(X^2 + \gamma^2 Y^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right) \quad \cdot$$
$$X = x\cos\theta + y\sin\theta \qquad Y = -x\sin\theta + y\cos\theta \quad \cdot \tag{1}$$

where, orientation θ, the effective width σ, the wavelength λ, the aspect ratio $\gamma = 0.3$. In this paper, 9 bands, 2 scales in each band, and 4 orientations (90°, -45°, 0°, 45°) are adopted. The values of these parameters are set based on [13]. Given an image, each pixel is convoluted with Gabor filters, resulting in a series of Gabor images with expressional features (e.g. bar and edge).

2.2 $(2D)^2$PCA

PCA-based methods have been widely used for dimension reduction, however, most of the methods need to reshape 2D image into a 1D feature vector, which leads to three problems: the intrinsic 2D structure of an image is removed, curse of dimensionality dilemma and small sample size [14]. Thus, $(2D)^2$PCA [15] was used in our framework to directly calculate the feature without matrix-to-vector conversion, and save storage requirement by performing PCA on row and column pixels simultaneously to obtain feature matrixes that represent images.

2.3 Layer Selection

The tasks of selecting 'salient' matrix layers and determining their optimized sizes are completed by using the recognition performance of an image set from the JAFFE database. The test set includes images with the emotion index '1', whilst the training set comprises of the rest images. As for optimized sizes, a size range [4, 40] with an interval of 2 is chosen based on preliminary experiments. The selection process can be described as follows.

Let L_{bsot} be the b^{th} band, s^{th} scale and o^{th} orientation layer of training image A_t (b = 1,2,…, 9; s = 1,2; o = 1,2,3,4; t = 1,2,…, M, M is the number of training images), the feature matrix of L_{bsot} is F_{bsot}. Let L_{bsol} be the b^{th} band, s^{th} scale and o^{th} orientation layer of test image T_l (l = 1,2,…, Q; Q is the number of test images), the feature matrix of L_{bsol} is F_{bsol}. The distance between L_{bsot} and L_{bsol} is defined by

$$D(L_{bsot}, L_{bsol}) = \left\| F_{bsot} - F_{bsol} \right\| \quad . \tag{2}$$

where $\|\cdot\|$ is the L1 or L2 norm of ($F_{bsot} - F_{bsol}$).

Then, the correct recognition rate (CRR) of the b^{th} band, s^{th} scale and o^{th} orientation layer of all test images can be obtained by the nearest neighbor classifier using these distances. Based on the results, layers with comparatively higher CRRs are selected as 'salient' layers. For each 'salient' layer, the optimized size is set to be a little bigger than the size of the best performance in order to gain a general performance. Finally, a total of 26 'salient' layers and their optimized sizes are obtained and listed in Table 1, in which BSO '322' represents the 3th band, 2th scale and 2th orientation, L2 stands for using L2 distance.

Table 1. The selected 'salient' feature matrix layers and sizes

BSO	Size	BSO	Size	BSO	Size	BSO	Size	BSO	Size
322	20	513	22	622(L2)	18	723	16	913	16
412	20	522	16	623	10	724	14	923	14
413	20	523	14	624	16	812	10	-	-
422	18	612(L2)	18	712	12	813	16	-	-
423	18	613	20	713	18	814	22	-	-
512	18	614	16	714	14	823	14	-	-

2.4 Layer Fusion

The layer fusion step performs FER by fusing the 'salient' feature matrix layers with optimized sizes. Firstly, for each 'salient' layer of one test image, KNN is used to calculate the K possible expressions. Then the expressions of all layers are combined to obtain the final result using the maximum rule. The algorithm is as follows.

Let L_{pt} be the p^{th} layer of training image A_t (p = 1,2,…, 26; t = 1,2,…, M), and L_{pl} be the p^{th} 'salient' layer of test image T_l (l = 1,2,…, Q), their feature matrixes are F_{pt} and F_{pl} respectively. For each L_{pl}, the M distances D(L_{pt}, L_{pl}) between L_{pl} and L_{pt} of all training images can be calculated by the equation (2). The nearest distance of the M distances is defined by

$$D(L_{pt}, L_{pl}) = \min_{t=1}^{M} \left\| F_{pt} - F_{pl} \right\| \quad . \tag{3}$$

Similarly, the K smallest D(L_{pt}, L_{pl}) also can be obtained, the emotion labels E_{pi}^{g} (i = 1,2,…,K; g = 1,2,…,6; $E_{pi}^{g} \in \{AN, DI, FE, HA, SA, SU\}$) of these chosen K L_{pk} are recorded. Then E_{pi} with the same emotion label are summed over 26 'salient' layers:

$$E_g = \max_{g=1}^{6} \left(\sum_{p=1}^{26} \sum_{i=1}^{k} E_{pi}^{g} \right) \quad . \tag{4}$$

Thus, the final output of emotion g corresponds to the largest E_g.

3 Experiments

3.1 Databases

The JAFFE database [4] contains 213 gray images of 7 facial expressions posed by 10 Japanese females. Each object has 3 or 4 frontal face images for each expression. The name of each image is identified by subject name initials, emotion initials & index, and image index. Cohn-Kanade database [16] includes 2105 image sequences from 182 subjects ranged in age from 18 to 30 years. Image sequences were digitized from neutral to target display. The six universal expressions were based on descriptions of prototypic emotions. In this paper, all the images of the six universal expressions from JAFFE are used. For CK, 1184 images that represent one of the six expressions are selected, 4 images for each expression of totally 92 subjects. The images are chosen from the last image of each sequence, then one every two images. The faces of all images from two databases are cropped and scaled to a resolution of 110*110 pixels.

3.2 JAFFE Database Tests

For each validation step, the images with the same emotion index are grouped as the test set, and the remaining images are regarded as the training set. In this research, only emotion index from 1 to 3 are tested due to the fact that most subjects do not contain images with emotion index '4'. As a test benchmark, all layer (AL) approach is defined as using all layer features and L1 distance.

The CRRs of three test sets using KNN with K ranged from 1 to 14 are shown in Fig. 2. As shown in this figure, for both the proposed and AL approaches, the highest CRR of each set is obtained by KKN when K=1. Regarding the highest CRR of each set, the proposed approach shares the same value (90.0%) with AL approach in set1 and achieves bigger values than all layer approach in set2 and set3. The highest CRR (96.923%) of the proposed approach is 3.077% bigger than that of the AL approach. Therefore, it can be concluded that the chosen and optimized layers can achieve better recognition performance than using all layers. Since the training and test images of set2 and set3 are different from those used for obtaining the chosen and optimized

layers, their high performances indicate a good general recognition capability of these 'salient' layers.

Among the three sets, set2 obtains the best overall recognition performance for all K values and keeps the highest CRRs in both the proposed and AL approaches, while set1 ranks the lowest. After the peak performance in both approaches, the CRRs decrease as K values increase, whereas CRRs of the proposed approach decrease quicker than those of AL approach. The reason is probably that AL approach utilizes all expressional information for FER, thus a steady decline of CRR is expected, whereas the proposed approach only adopts part of this information, therefore, a rapid decrease is anticipated.

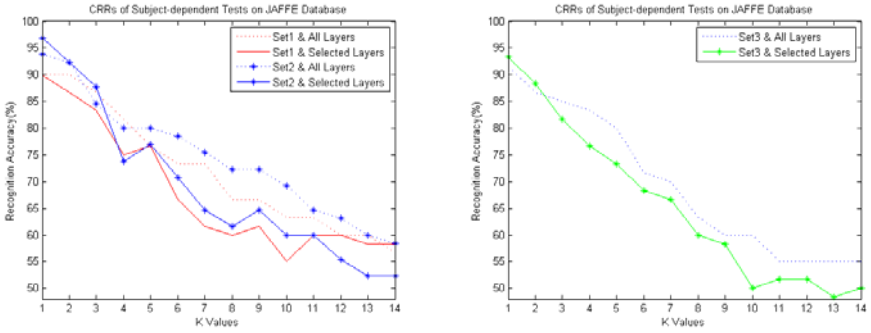

Fig. 2. CRR comparisons between the proposed and AL approaches using JAFFE database

The confusion matrix of the six expressions can be drawn by setting K to be 1 for all three sets in order to obtain the highest CRRs. The result is demonstrated in Table 2. As shown in this table, the images of surprise are all correctly recognized probably due to the apparent characteristic of big mouth; the second best recognized emotion is anger, and only one image is falsely classified as sad. On the other hand, happy is the most difficult emotion to distinguish from others. Another interesting point is that sad is the emotion that is most likely to be incorrectly recognized as target emotion. And this may be owing to the erratic expressers on sad in JAFFE, which is in accord with the work [5] that reported two erratic expressers (UY and NA) existed in JAFFE.

Table 2. Confusion matrix of six expressions using JAFFE database

	AN	DI	FE	HA	SA	SU	Overall
AN	29	0	0	0	1	0	96.7%
DI	0	28	1	0	0	1	93.3%
FE	0	1	27	0	2	0	90.0%
HA	0	0	0	26	4	0	86.7%
SA	0	0	1	1	28	0	93.3%
SU	0	0	0	0	0	30	100%

3.3 CK Database Tests

Since each subject has four images for each expression, all images can be classified into four sets that include one of the four images per set. Four cross-validation tests are conducted separately and the results are compared with the AL approach as shown in Fig. 3. Based on the graphs, the overall performances of the two approaches are fairly satisfactory. For all the four sets, both approaches achieve their highest CRRs when K=1 and 2, but the AL approach can retain the highest CRR of 100% with a big K value (for instance, 6 in set3). As for set1, set2 and set3, the highest CRRs of both approaches is 100%, while for set4, the highest CRR (99.662%) of the proposed approach is 0.338% lower than that of the AL approach since one happy image is wrongly recognized as fear. For both approaches, the CRRs decrease when K increases. However, similar to our findings while using JAFFE database, CRR of the proposed approach declines quicker than that of the AL approach. Thus, we can conclude that the selected 'salient' layers can achieve higher recognition performances compared to using all layers.

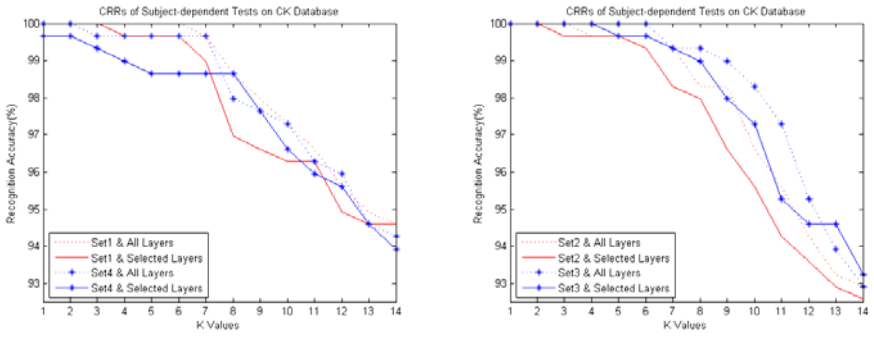

Fig. 3. CRR comparisons between the proposed and AL approaches using CK database

3.4 Computational Time Comparison

For each of JAFFE and CK, the average computational time of all test images at three stages, including Gabor feature, $(2D)^2$PCA and KNN, is calculated and demonstrated in Table 3. The program was developed by Matlab 7.0.1 under a laptop configuration of core duo 1.66GHz CUP and 2GB memory. Based on the time, the proposed approach has shown a substantial improvement compared with the AL approach as it has reduced 75% to 80% of the processing time in the AL approach. Moreover, there is an 75% to 82% of time reduction for computing Gabor features, 65% to 75% for processing $(2D)^2$PCA, and 75% to 80% for recognizing expressions using KNN. Time spent on computing Gabor features is nearly 90% of the overall time for JAFFE, and about 60% for CK. This demonstrates that Gabor feature is the most computationally expensive. On the other hand, $(2D)^2$PCA only requires 2.7% to 4.7% of the overall time. Another notable point is that the computational time of KNN on CK is 6 to 7

times as much as that on JAFFE. This is due to KNN has a bigger number of test and training images to process as CK contains more images than JAFFE.

Table 3. Computational time comparisons at three stages (in seconds)

	Proposed approach				All layer approach			
	Gabor	$(2D)^2PCA$	KNN	Total	Gabor	$(2D)^2PCA$	KNN	Total
JAFFE	0.301	0.016	0.025	0.342	1.263	0.047	0.101	1.411
CK	0.244	0.011	0.150	0.405	1.342	0.047	0.711	2.100

3.5 Comparisons with Previous Work

In this paper, the performances of Liang [17] (using LLE) and Guo [18] (using FSLP) are used as the benchmark for JAFFE, while the performances of Wang [19] (using NBC and QDC) and Wong [20] (using FEETS) are used as the benchmark for CK. The choice on these benchmarked works is based on the database images being the most similar to our work. The comparison results are shown in Table 4, from which we can see that using JAFFE, the proposed approach exceeds Liang's approach by 1.90% with respect to the maximum (Max) CRR, 0.6% for the average (Ave) CRR, and 4.3% for the minimum (Min) CRR. Moreover, it exceeds Guo's approach by 2.4% for the Ave CRR. A better performance is shown by the CK database as the proposed approach surpasses Wong's by 6.71% for the Max CRR and 17.14% for the Min CRR, while it also surpasses Wang's approach by 3.43% for the Max CRR, and 12.45% for the Min CRR. Hence, our experiment has demonstrated a significant recognition improvement in the proposed approach compared to the previous work.

Table 4. CRR comparisons with previous work (%)

	Proposed approach			[17] and [19]			[18] and [20]		
	Max	Ave	Min	Max	Ave	Min	Max	Ave	Min
JAFFE	96.9	93.4	90.0	95	92.8	85.7	-	91.0	-
CK	100	99.89	99.66	93.29	-	82.52	96.57	-	87.21

4 Conclusions

This paper presents a novel method to automatically select, optimize and fuse 'salient' Gabor layers to improve the current performance in FER from still images. The experiments on JAFFE and CK databases demonstrate that the proposed approach can achieve significant improvements on recognition performance and computational time compared to the previous work. Our results confirm that wider range of Gabor filters can improve the performance as expressional information is evenly distributed over these filters. Moreover, our experiments show that the time used for computing Gabor filters takes a large part of the overall processing time of our framework, while $(2D)^2PCA$ only requires a small proportion of the overall time.

In our future work, we aim to conduct more experiments to improve CRR by increasing orientation number. Meanwhile, the combination of $(2D)^2PCA$ with other local feature extraction methods (for example, local binary pattern [21]) seems to be a promising direction. Another important field is combining both appearance and motion features for FRE since researches [22] have confirmed the significant role of dynamic information in the process of expressing and recognizing facial expressions.

Acknowledgments. The authors would like to thank Nicki Ridgeway for providing the Cohn-Kanade AU-Coded Facial Expression Database and the providers of JAFFE database.

References

1. Deng, H.B., Jin, L.W., Zhen, L.X., Huang, J.C.: A new facial expression recognition method based on local gabor filter bank and pca plus lda. International Journal of Information Technology 11 (2005) 86-96
2. Caifeng, S., Shaogang, G., McOwan, P.W.: Robust facial expression recognition using local binary patterns. Image Processing, 2005. ICIP 2005. IEEE International Conference on, Vol. 2 (2005) II-370-373
3. Wei Feng, L., ZengFu, W.: Facial Expression Recognition Based on Fusion of Multiple Gabor Features. Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Vol. 3 (2006) 536-539
4. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on (1998) 200-205
5. Bashyal, S., Venayagamoorthy, G.K.: Recognition of facial expressions using Gabor wavelets and learning vector quantization. Engineering Applications of Artificial Intelligence 21 (2008) 1056-1064
6. Chen, H.Y., Huang, C.L., Fu, C.M.: Hybrid-boost learning for multi-pose face detection and facial expression recognition. Pattern Recognition 41 (2008) 1173-1185
7. Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J.: Dynamics of facial expression extracted automatically from video. Image and Vision Computing 24 (2006) 615-625
8. Yu, J., Bhanu, B.: Evolutionary feature synthesis for facial expression recognition. Pattern Recognition Letters 27 (2006) 1289-1298
9. Gunes, T., Polat, E.: Feature selection for multi-SVM classifiers in facial expression classification. Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on (2008) 1-5
10. Lajevardi, S.M., Lech, M.: Facial Expression Recognition Using Neural Networks and Log-Gabor Filters. Computing: Techniques and Applications, 2008. DICTA '08.Digital Image (2008) 77-83
11. Kong, A.: An evaluation of Gabor orientation as a feature for face recognition. Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (2008) 1-4
12. Dadgostar, M., Tabrizi, P.R., Fatemizadeh, E., Soltanian-Zadeh, H.: Feature Extraction Using Gabor-Filter and Recursive Fisher Linear Discriminant with Application in Fingerprint Identification. Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on (2009) 217-220

13. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust Object Recognition with Cortex-Like Mechanisms. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (2007) 411-426
14. Kong, H., Wang, L., Teoh, E.K., Li, X., Wang, J.-G., Venkateswarlu, R.: Generalized 2D principal component analysis for face image representation and recognition. Neural Networks 18 (2005) 585-594
15. Zhang, D., Zhou, Z.-H.: $(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing 69 (2005) 224-231
16. Kanade, T., Cohn, J.F., Yingli, T.: Comprehensive database for facial expression analysis. Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on (2000) 46-53
17. Liang, D., Yang, J., Zheng, Z., Chang, Y.: A facial expression recognition system based on supervised locally linear embedding. Pattern Recognition Letters 26 (2005) 2374-2389
18. Guo, G., Dyer, C.R.: Learning from examples in the small sample case: face expression recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part B 35 (2005) 477-488
19. Wang, J., Yin, L.: Static topographic modeling for facial expression recognition and analysis. Comput. Vis. Image Underst. 108 19-34
20. Wong, J.-J., Cho, S.-Y.: A face emotion tree structure representation with probabilistic recursive neural network modeling. Neural Computing & Applications (2008)
21. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing 27 (2009) 803-816
22. Yongmian, Z., Qiang, J., Zhiwei, Z., Beifang, Y.: Dynamic Facial Expression Analysis and Synthesis With MPEG-4 Facial Animation Parameters. Circuits and Systems for Video Technology, IEEE Transactions on 18 (2008) 1383-1396