Evaluation of Smile Detection Methods with Images in Real-world Scenarios

Zhoucong Cui, Shuo Zhang, Jiani Hu, Weihong Deng

Beijing University of Posts and Telecommunications, Beijing, China

Abstract. Discriminative methods such as SVM, have been validated extremely efficient in pattern recognition issues. We present a systematic study on smile detection with different SVM classifiers. We experimented with linear SVM classifier, RBF kernel SVM classifier and a recently-proposed local linear SVM (LL-SVM) classifier. In this paper, we focus on smile detection in face images captured in real-world scenarios, such as those in GENKI4K database. In the meantime, illumination normalization, alignment and feature representation methods are also taken into consideration. Compared with the commonly used pixel-based representation, we find that local-feature-based methods achieve not only higher detection performance but also better robustness against misalignment. Almost all the illumination normalization methods have no effect on the detection accuracy. Among all the SVM classifiers, the novel LL-SVM is verified to find a balance between accuracy and efficiency. And among all the features including pixel value intensity, Gabor, LBP and HOG features, we find that HOG features are the most appropriate features to detect smiling faces, which, combined with RBF kernel SVM, achieve an accuracy of 93.25% on GENKI4K database.

1 Introduction

Facial expression recognition has been an active topic for the last two decades. Common methods such as feature-point-based expression classifiers [1] [2] [3], 3D face modeling [4] [5] and dynamic analysis of video sequences [6] [7] [8] have achieved great success, while few have specially focused on smile detection. Shinohara and Otsu [9] used a Fisher weight map as an assist to gain an accuracy of 97.9% on 96 testing face images. The result is desirable, but the data is limited. Another smile detector presented in [10] achieves an accuracy of 96.1% on 4928 testing face images. However, the face images used in these studies are captured in limited conditions, and are mainly frontal. Many of the testing images were collected by asking subjects to deliberately pose certain expressions, thus exaggerating the effect of the expressions, which would seldom occur in our daily life. In this paper, we focus on face images collected in real-world scenarios, and present a systematic study on smile detection with different SVM classifiers using various features. To measure the real-world performance on smile detection, we choose the GENKI4K database and some of the image examples are shown in Fig 1.

As regards classifiers, support vector machine has been universally accepted as an outstanding classifier for both of its efficiency and accuracy. Most SVM-based works adopt RBF kernel SVM, but one point deserves mentioning is that as the size of the training data increases, RBF is much more time-consuming compared to linear SVM.

Fig. 1. Examples of real-life faces from the GENKI4K database. The top two rows are smile faces and the rest are nonsmile faces.

Another classifier, Orthogonal Coordinate Coding (OCC) SVM, seems to strike a balance between accuracy and efficiency, which we will discuss later.

This paper is organized as follows. In Section 2, we will give a brief introduction to the features we used. Then some SVM classifiers will be discussed in Section 3, including the improved OCC. We also provide a sensitivity analysis in Section 4 and Section 5 will demonstrate the conclusions.

2 Features

This section introduces three well-known local features for image description, which have been evaluated in previous research on face recognition [11][12][13]. We would like to evaluate and compare their performance on smile detection tasks.

2.1 Local Binary Pattern

The original LBP operator, introduced by Ojalaet al.[14], is a powerful means of texture description. The operator labels the pixels of an image by thresholding the 3×3 neighbourhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels are used as a texture descriptor.

Later research extended the operator to use neighbourhood of different sizes [15]. Using circular neighbourhood and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighbourhood.

Rotation invariance was later taken into consideration in LBP-based representation due to the circular property of the pattern. The experiment done by Ojala et al. (Ojala et al., 2002) indicates that only a particular subset of local binary patterns (those containing at most two bitwise transitions from 0 to 1 or vice versa) are typically present in most of the pixels contained in real images, and these patterns are referred to as uniform [15].

Concretely, the LBP operator is employed to obtain the LBP value of each pixel of the image in the uniform mode. Then the whole image is divided into same-sized small blocks of which a histogram is calculated, containing information about the distribution of the local texture, such as edges, spots and flat areas. Finally, regional histograms are concatenated to build a global description of the face image.

2.2 Histogram of Oriented Gradient

The basic idea behind the Histogram of Oriented Gradient descriptors is that local appearance and shapes within an image can be described by the distribution of intensity gradients or edge directions. The implementation of these descriptors can be achieved by dividing the image into small connected regions, called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then forms a descriptor. To improve accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. The normalization results in better invariance to changes in illumination or shadowing.

With normalized pixel values, we calculate the gradient of each pixel point. Then we divide the face image into small cells, for example, 5*5, and calculate the histogram of each cell to form the cell descriptor. And we combine several cells to a block, and concatenate each cell descriptor into a block descriptor. Finally, block descriptors are concatenated to build a global description of the face.

2.3 Gabor Features

A Gabor filter can be seen as a sinusoidal plane of a particular frequency and orientation, modulated by a Gaussian envelop [16][17]. A 2-D Gabor function g(x, y) and its Fourier transform G(u, v) are defined as

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp[-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + (\frac{y^2}{\sigma_y^2}) + 2\pi jWx] \tag{1}$$

where $j = \sqrt{-1}$, and W is the frequency of the modulated sinusoid

$$G(u,v) = exp[-\frac{1}{2}[\frac{u - W^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}] \tag{2}$$

where $\sigma_u = 1/2\pi\sigma_x, \sigma_v = 1/2\pi\sigma_y$

A self-similar filter dictionary can be obtained by associating an appropriate scale factor α and a rotation parameter θ with the mother wavelet g(x,y) . M and N represent the scales and orientations of the Gabor wavelets, respectively.

$$g_{mn}(x,y) = \alpha(x^{'}, y^{'}), 0 \leq m \leq M-1, 0 \leq n \leq N-1 \tag{3}$$

where
$x^{'} = \alpha^{-m}(xcos\theta + ysin\theta), y^{'} = \alpha^{-m}(-xsin\theta + ycos\theta)$ and $\theta = n\pi/K$, while K is the total number of orientations.

3 SVM

As a supervised learning model, the well-known support vector machine has already gained its dominance in classification issues. An interesting property of SVM is that it is an approximate implementation of the Structural Risk Minimisation (SRM) induction principle that aims at minimising a bound on the generalisation error of a model, rather than minimising the mean square error over the data set.

The original SVM was proposed to solve binary classification problems by finding the optimal separating hyperplane. It has achieved great success in linear separable problems. However, most practical problems are not linearly separable. To cope with this problem, previous research has found solutions by using kernel tricks, which maps the data to a higher dimension. The kernel function effectively solves the nonlinear problems, but in the meantime, it also brings limitations in practice: First, the complexity is highly dependent on the size of the training data, which means as the size of the training set grows, it takes increasingly long time to test a smile. Another limitation is the lack of theoretic support on how to choose kernel functions and setting parameters for specific problems.

Recently, a novel locally linear SVM has been proposed, which has smooth decision boundary and bounded curvature.

The standard linear SVM classifier is adopted as our first classifier, which takes the form

$$H(x) = \sum_{i=1}^{m} \alpha_i y^{(i)} < x^{(i)}, x > +b \tag{4}$$

where $H(x) = 0$ is the boundary plane, $x^{((i))}$is the support vector and $y^{(i)}$, α_i is the responding label and weight respectively. b is the bias value.

As for kernel SVM, it takes the form

$$H(x) = \sum_{j=1} y_j \alpha_j K(x, x^{(j)}) + b \tag{5}$$

where $K(x, x_j)$ stands for the kernel we choose. We choose an SVM classifier with square kernel as our second classifier, which is $K(x, z) = (x^T z)^2$. For our third classifier, we adopt an RBF kernel, which is

$$K(x, z) = e^{-\frac{||x-z||^2}{2\sigma^2}} \tag{6}$$

To introduce the orthogonal coordinate coding, first we take another look at the standard linear SVM classifier function:

$$H(x) = w(x)^T x + b(x) \tag{7}$$

The thought behind locally linear SVM is that w is adapted with different input x, which could be described as

$$H(x) \approx \sum (c_v w(v)^T x + c_v b) =< W^T, c_x x^T > +c_x^2 b \tag{8}$$

where v belongs to the coding anchor points and $c_v(x)$ is the related coefficient. Imagine we select m anchor points, then W^T would be an m by n matrix and each row $w(v)^T$is the corresponding anchor vector.

To preserve local characteristics, large amounts of anchor points are needed, which plays a significant role in the performance of locally linear SVM. As a result, if the input data is sparse, it could lead to many holes in the feature space. To solve this problem, Ziming Zhang [18] proposed a new LL-SVM exploiting an orthogonal coordinate coding (OCC) scheme, which encodes data using orthogonal anchor planes rather than anchor points.

Given an input matrix $X = [x^1, x^2, ...]$, if we select N orthogonal vectors to code X, the algorithm is presented as follows

Algorithm 1 The OCC coding algorithm.

Require: X: data matrix,$X = [x^1, x^2, ...]$
Ensure: find the coding vector$\widehat{c}_x$
1: Calculate the SVD of X . $X = U\Sigma V$, where U and V are two unitary matrices. Σ is a rectangular diagonal matrix, of which the elements are all non-negative real numbers and set in descending order;
2: We select a few top eigenvectors u_i as our orthogonal basis vectors to form the basic matrix $\widehat{U}$. Consider one input x, we code it as $\widehat{c}_i = \frac{x^T u_i}{\sigma_i}$, where σ_i is the corresponding singular value. Then we define a generator matrix $G = \widehat{U}\Sigma^{-1}$ and $\Sigma = \begin{bmatrix} \sigma_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & \sigma_n \end{bmatrix}$. As a result, given a vector x, we code it as $\widehat{c}_x = G^T x$;
3: Normalize the $\widehat{c}_x$ as $\widehat{c}_x = \widehat{c}_x = \frac{\widehat{c}_x}{||\widehat{c}_x||_1}$;

4 Experiment

In this section we analyze the effect of illumination normalization, face image scales, face alignment and various controlled misalignments on smile detection.

4.1 Data

By far, almost all the expression recognition studies have been carried out on facial expression databases that were collected under tightly controlled conditions containing limited number of subjects.[19] [20]. Hence, it leads to a lack of diversity in illumination conditions and individual differences which is unavoidable in our real world. To address this problem, we carried out our experiments on the publicly-available GENKI4K database. The face images were taken not by laboratory scientists, but by common people from all over the world taking photographs of themselves for personal purpose. The database consists of 4000 face images, of which 2162 are smile faces, and 1838 are nonsmile faces as shown in Fig 1. As we can see, the images vary significantly in illumination conditions as well as poses and resolutions.

Fig. 2. (a) Face images that are detected directly by the detector provided by Open CV 1.0. (b) Face images that are aligned and cropped based on the manually labeled eyes positions

In our experiment, the images were first converted to gray scale. As for scaling and alignment, one set of images are normalized to into canonical faces of three different sizes: 24×24, 36×36 and 48×48 pixels, which is done using the face detector provided by Open CV 1.0. Another set of images are normalized to the same three levels but based on the manually labeled eye positions. Fig 2 illustrates some examples of the aligned face images and the unaligned ones.

4.2 Procedure

As regards feature extraction in the training and testing processes, we choose pixel value intensity, LBP, HOG, and Gabor features. Three classifiers, including linear SVM, OCC SVM and RBF kernel SVM, were applied in the experiments. We divide the face images into 4 similar sets and each contains similar number of smile faces and nonsmile faces. All of the sets are applied to a fourfold cross-validation, which means when one set is used as the test data, the other three sets are used as training data. The procedure repeats four times for each set. Test variables are shown in Table 1.

A. Illumination Normalization

Face images from Fig. 1 and Fig. 2 witness a wide range of illumination conditions. To alleviate the impact of illumination conditions, we adopt several illumination normalization methods proposed in previous research for our experiment:

1) Gaussian disposition : It first filters out the uncontrolled illumination changes by dividing the intensity value of each pixel in the original image by the one after performing Gaussian Smooth [21].

2) Histogram equalization (HE): HE is a simple and widely-used technique for normalizing illumination effects.

3) Discrete cosine transform (DCT): The DCT-based normalization [22] sets a number of DCT coefficients corresponding to low frequencies as zero to achieve illumination invariance.

Table 1. Test Variables and Their Conditions.

Variable	Conditions
Features	Pixel value intensity LBP features Gabor features HOG features
Classifier	Linear SVM OCC RBF SVM
Alignment	None Manual
Input image size	24×24 36×36 48×48

We apply each of these methods to grayscale face images and then conduct face detection with the four features and three SVM classifiers mentioned above. The results can be seen in Table 2. Abnormally, it seems all the illumination normalization methods fail to work except the case of HE method employed with OCC and RBF SVM classifiers using LBP feature. All the other methods with the four features, however, bring the recognition rate down. It might be caused by the complexity of illumination conditions in the real world. Another possible explanation could be the characteristics of a smile face mainly rely on the structure of the face itself and the state of the organs, especially the mouth. Since these illumination normalization methods do not work as expected, the following experiments omit the illumination normalization step.

Table 2. Experiment Results of Different Illumination Normalization Methods.

Illumination Normalization	Features	Linear(%)	OCC(%)	RBF SVM(%)
None	Pixel	89.70	91.80	91.50
	LBP	92.28	92.47	92.38
	Gabor	91.13	92.03	92.08
	HOG	93.05	93.15	93.25
Gaussian	Pixel	84.70	87.35	87.35
	LBP	89.45	89.25	89.25
	Gabor	89.20	89.42	90.10
	HOG	91.70	92.03	92.13
HE	Pixel	88.50	91.30	90.65
	LBP	92.15	**92.70**	**92.40**
	Gabor	89.95	90.35	91.13
	HOG	91.15	92.48	92.00
DCT	Pixel	88.10	90.52	90.57
	LBP	91.32	91.47	91.52
	Gabor	90.56	91.05	90.89
	HOG	91.03	92.30	92.40

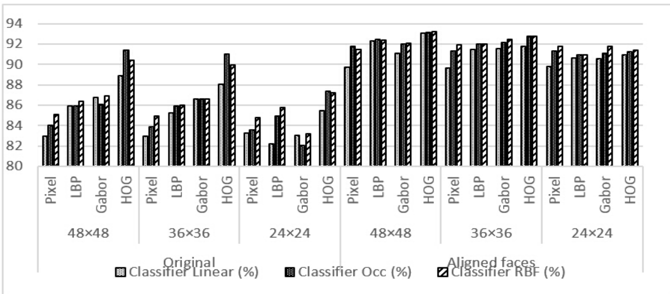


Fig. 3. Classification rates with different classifiers and face sizes

B. Scales and Alignment

The alignment method is conducted as follows: Firstly, we manually find the eye locations of each face image. Then we rotate the images to make the two eyes in the same horizontal line. Finally, we scale and crop the face images to ensure the two eyes to be at fixed positions.

All the methods achieve better classification rate with manually aligned faces than with unaligned ones, as can be seen from Table 3. Besides, it can be noticed that with unaligned face images, small-size images (24×24) are inferior to the larger ones in terms of classification rate. In contrast, as the size varies, the rate

Table 3. Classification Rates with Different Classifiers and Face Sizes.

	Image size	**Feature**	**Classifier**		
			Linear (%)	**OCC(%)**	**RBF(%)**
Original faces	48×48	Pixel	82.98	84.06	85.09
		LBP	85.93	85.96	86.39
		Gabor	86.73	86.07	86.91
		HOG	88.87	91.40	90.40
	36×36	Pixel	82.98	83.91	84.91
		LBP	85.22	85.91	86.01
		Gabor	86.63	86.65	86.65
		HOG	88.07	91.00	89.95
	24×24	Pixel	83.26	83.59	84.75
		LBP	82.19	84.94	85.80
		Gabor	83.04	82.05	83.22
		HOG	85.44	87.36	87.20
Aligned faces	48×48	Pixel	89.70	91.80	91.50
		LBP	92.28	92.47	92.38
		Gabor	91.13	92.03	92.08
		HOG	93.05	93.15	93.25
	36×36	Pixel	89.63	91.36	91.95
		LBP	91.50	92.00	92.03
		Gabor	91.52	92.13	92.48
		HOG	91.80	92.78	92.80
	24×24	Pixel	89.83	91.35	91.80
		LBP	90.63	90.93	90.93
		Gabor	90.55	91.08	83.22
		HOG	90.98	91.28	91.43

with aligned faces seems much more robust.

The classification rates for different face image sizes are illustrated in Table 3 and Fig 3. Almost all the best classification rates are achieved with the image size 48×48, but compared with the size of 36×36, there is no statistically significant difference. When the image size goes down to 24×24, almost all the classification rates have a more noticeable drop.

It can also be observed that, regardless of the image size and the feature chosen, the OCC and RBF SVM classifiers perform almost equally well, much better than the linear SVM classifier does.

As for features, all the other features outperform the pixel value intensity. HOG feature performs the best almost under any condition, especially with the original faces of 48×48 and 36×36 sizes. The classification rates with LBP and Gabor features achieve similar results. However, with the size of 24×24, the advantage of the other features over pixel value intensity becomes less obvious. One possible reason is that when the image size gets smaller, local texture could not be represented by the LBP and Gabor features appropriately.

C. Sensitivity Analysis

In this section, we give an analysis on the sensitivity of smile classifiers to various detection and alignment inaccuracies as Baluja and Rowley did in [23]. We use the manually aligned faces and add translations in horizontal and vertical directions from -3 to +3 pixels, in-plane rotation of the faces from $-45°$ to $+45°$, and Gaussian noise with the variance from 1 to 7.

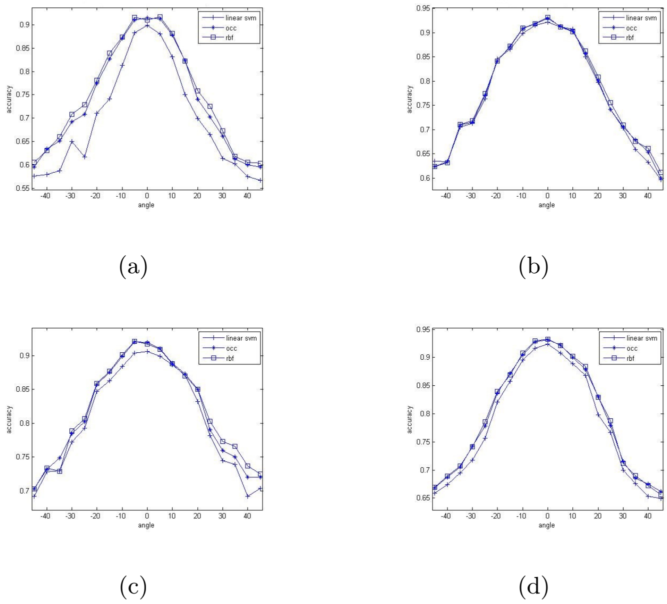


Fig. 4. Sensitivity against rotation. (a), (b), (c), (d) illustrate the sensitivity against rotation with features of pixel intensity, LBP, Gabor and HOG respectively.

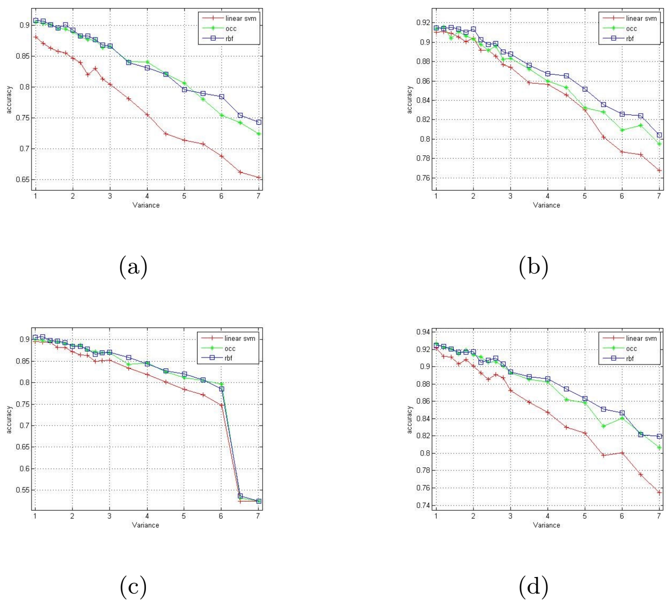
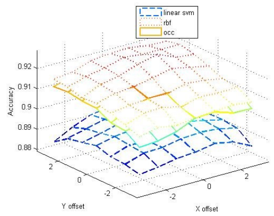
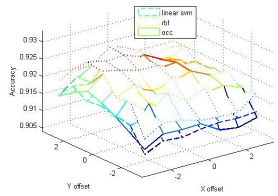


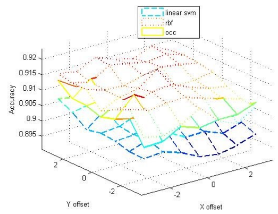
Fig. 6. Sensitivity of classifiers against Gaussian noise (a), (b), (c), (d) illustrate the sensitivity against Gaussian noise with features of pixel intensity, LBP, Gabor and HOG respectively.



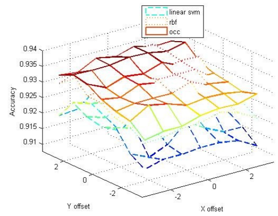
(a)



(b)



(c)



(d)

Fig. 5. Sensitivity of the classifiers against translation. (a), (b), (c), (d) illustrate the sensitivity against translation with features of pixel intensity, LBP, Gabor and HOG respectively.

As expected, we observe significant advantages of the three other features over pixel value intensity for they are more robust against image rotation, translation and Gaussian noise.

From the sensitivity analysis against rotation shown in Fig. 4, we can see that Gabor feature shows the lowest sensitivity against rotation, since its recognition rate remains the highest among the four features under the same conditions.

As for sensitivity against translation, which is shown in Fig. 5, HOG feature turns out to be the best among all the features.

As shown in Fig. 6, with the variance of Gaussian noise increasing from 1 to 7, the average classification rate with pixel value intensity decrease from 90.7% to 73.1%, LBP feature from 91.3% to 78.9% and HOG feature from 92.43% to 79.38%. Moreover, we observe that with the variance of Gaussian noise grows, the OCC and RBF SVM classifiers depict increasingly notable advantage over the linear SVM classifier.

Besides, in almost all the sensitivity analyses above, the OCC classifier perform equally well with RBF SVM classifier, outperforming the linear SVM classifier.

5 Summary and Conclusions

Database: Most of the existing facial expression databases were collected under strictly controlled conditions containing limited number of subjects, which lack diversity in illumination conditions and individual differences. To address this problem, we select the publicly-available GENKI4K database which was built with world-wide self-portrait photos under various conditions.

Illumination and size: As the results elucidated, there is no method found to fit the real-world scenarios, for the tested illumination normalization methods do not work properly to increase the smile detection rates and sometimes even prove to lower the rates. It might be caused by the diversity of the real-world environment. The size of input face images exerts limited impact on the smile detection rates. Almost all the best classification rates are achieved with the image size of 48×48, but the rates given by the size of 36×36 present no statistically difference. In general, the size of the images does not make much difference, but if it is too small, it may affect some of the features such as LBP and Gabor.

Alignment: As expected, the recognition rate is improved obviously with the alignment method. Under the same conditions including size, features and classifiers, the aligned face images all gain a better result. Surprisingly, under the same settings except for the alignment, on average, our result is 5.88 % higher than Shan and Caifeng's method in [24]. Our alignment method proves to be very effective. Besides, it is very simple and easy to conduct.

Features and Classifiers: Among the features of pixel value intensity, LBP, Gabor and HOG features, HOG feature can best represent a smiling face. In all the cases tested, it always gains the highest accuracy. All the other three features perform better than pixel value intensity, but one point deserves mentioning is that when the size of face images is small, LBP and Gabor features tend to be inferior to pixel value intensity. As for the classifiers, in terms of accuracy, OCC and RBF SVM classifiers perform almost equally well and both outperform the linear SVM with pixel value intensity. And when it comes to time consumption, OCC SVM classifiers are ten times faster than the RBF kernel. Compared with linear SVM and RBF SVM, it seems that OCC SVM balances well between accuracy and efficiency.

Acknowledgement.

This work was partially sponsored by National Natural Science Foundation of China (NSFC) under Grant No. 61375031, No. 61471048, and No. 61273217. This work was also supported by the Fundamental Research Funds for the Central Universities, Beijing Higher Education Young Elite Teacher Project, and the Program for New Century Excellent Talents in University.

References

1. Pantic, M., Rothkrantz, L.J.: Facial action recognition for facial expression analysis from static face images. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **34** (2004) 1449–1461
2. Tian, Y.l., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on **23** (2001) 97–115
3. Kapoor, A., Qi, Y., Picard, R.W.: Fully automatic upper facial action recognition. In: Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on, IEEE (2003) 195–202
4. Wen, Z., Huang, T.S.: Capturing subtle facial motions in 3d face tracking. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, IEEE (2003) 1343–1350
5. Sebe, N., Sun, Y., Bakker, E., Lew, M.S., Cohen, I., Huang, T.S.: Towards authentic emotion recognition. In: Systems, Man and Cybernetics, 2004 IEEE International Conference on. Volume 1., IEEE (2004) 623–628
6. Cohn, J.F., Schmidt, K.L.: The timing of facial motion in posed and spontaneous smiles. International Journal of Wavelets, Multiresolution and Information Processing **2** (2004) 121–132
7. Cohen, I., Sebe, N., Garg, A., Chen, L.S., Huang, T.S.: Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and Image Understanding **91** (2003) 160–187
8. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. Pattern Analysis and Machine Intelligence, IEEE Transactions on **27** (2005) 699–714
9. Shinohara, Y., Otsu, N.: Facial expression recognition using fisher weight maps. In: Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, IEEE (2004) 499–504

10. Déniz, O., Castrillon, M., Lorenzo, J., Anton, L., Bueno, G.: Smile detection for user interfaces. In: Advances in Visual Computing. Springer (2008) 602–611
11. Deng, W., Hu, J., Guo, J.: Extended src: Undersampled face recognition via intraclass variant dictionary. IEEE Trans. Pattern Anal. Mach. Intell. **34** (2012) 1864–1870
12. Deng, W., Hu, J., Lu, J., Guo, J.: Transform-invariant pca: A unified approach to fully automatic face alignment, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **36** (2014) 1275–1284
13. Deng, W., Hu, J., Guo, J.: Linear ranking analysis. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2014)
14. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 971–987
15. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Computer vision-eccv 2004. Springer (2004) 469–481
16. Deng, W., Hu, J., Guo, J., Cai, W., Feng, D.: Emulating biological strategies for uncontrolled face recognition. Pattern Recognition **43** (2010) 2210–2223
17. Deng, W., Hu, J., Guo, J., Cai, W., Feng, D.: Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. Pattern Recognition **43** (2010) 1748–1762
18. Zhang, Z., Ladicky, L., Torr, P., Saffari, A.: Learning anchor planes for classification. In: Advances in Neural Information Processing Systems. (2011) 1611–1619
19. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, IEEE (2000) 46–53
20. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression database. Pattern Analysis and Machine Intelligence, IEEE Transactions on **25** (2003) 1615–1618
21. Wang, H., Li, S.Z., Wang, Y.: Generalized quotient image. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Volume 2., IEEE (2004) II–498
22. Chen, W., Er, M.J., Wu, S.: Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **36** (2006) 458–466
23. Baluja, S., Rowley, H.A.: Boosting sex identification performance. International Journal of Computer Vision **71** (2007) 111–119
24. Shan, C.: Smile detection by boosting pixel differences. Image Processing, IEEE Transactions on **21** (2012) 431–436