

# Facial Expression Recognition Using New Feature Extraction Algorithm

Hung-Fu Huang\* and Shen-Chuan Tai

Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan.

Received 10th Oct. 2011; accepted 5th Sep. 2012

---

## Abstract

This paper proposes a method for facial expression recognition. Facial feature vectors are generated from keypoint descriptors using Speeded-Up Robust Features. Each facial feature vector is then normalized and next the probability density function descriptor is generated. The distance between two probability density function descriptors is calculated using Kullback Leibler divergence. Mathematical equation is employed to select certain practicable probability density function descriptors for each grid, which are used as the initial classification. Subsequently, the corresponding weight of the class for each grid is determined using a weighted majority voting classifier. The class with the largest weight is output as the recognition result. The proposed method shows excellent performance when applied to the Japanese Female Facial Expression database.

**Keywords:** *Speeded-Up Robust Features, probability density function, Kullback Leibler divergence, weighted majority voting.*

---

## 1. Introduction

Facial expressions are important in communication. Mehrabian [1] found that in face-to-face communication of human, 7% of the communication information is transferred by linguistic language, 38% by paralanguage, and 55% by facial expressions.

Automating facial expression analysis and synthesis would allow the use of facial expressions in man-machine interaction. Facial expression recognition from static images is more difficult than that from image sequences due to the former having less information. Numerous algorithms for facial expression analysis from static images have been proposed [2].

The facial expression recognition process can be separated into feature extraction and classification based on pattern recognition theory. Many methods have been proposed, including principal component analysis [3], linear discriminant analysis [4], nonparametric discriminant analysis [5], optical flow [6-8], fisher weight maps [9], and local binary pattern [10].

Given images of a given object, feature detection and matching algorithms try to repeatedly detect the same point of interest in every image, regardless of the scale and orientation of the object, and match each point of interest from one image with the corresponding point in another image. The Scale Invariant Feature Transform (SIFT) [11-14]

---

Correspondence to: <hfhuang@seed.net.tw>

Recommended for acceptance by <Chulhee Lee>

ELCVIA ISSN: 1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

and the Speeded-Up Robust Features (SURF) [15-18] are widely used for this task. SURF has more discriminative power than SIFT.

The proposed method captures keypoint descriptors (i.e., feature vectors) from a facial expression image using SURF. Each feature vector is normalized to the unit length, creating a probability density function (PDF) descriptor. The Kullback Leibler (KL) divergence [19, 20] is applied to select the PDF descriptor. An equation is utilized to calculate and extract important PDF descriptors. The weighted majority voting (WMV) [21-23] classifier is adopted to calculate the matching scores of the respective localization regions. The recognition procedure is shown in Fig. 1. The Japanese Female Facial Expression (JAFFE) database was used in the experiments.

The paper is organized as follows. Section 2 describes the generation of keypoint features using the SURF method. The feature extraction from PDF is described in Section 3. Section 4 presents classification using WMV. The results are discussed in Section 5 and the conclusions are given in Section 6.

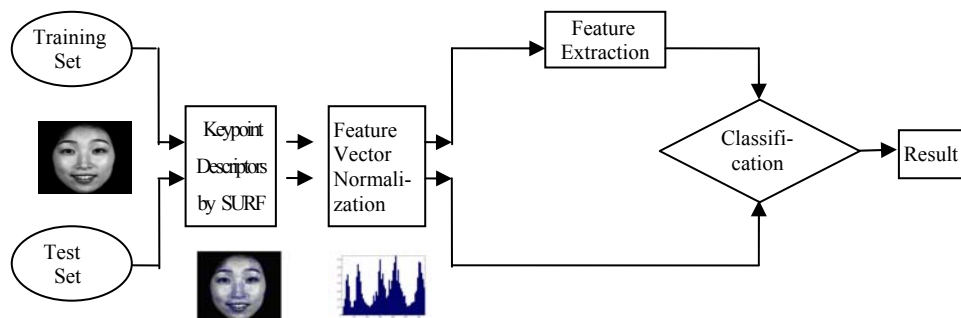


Fig.1 Facial expression recognition procedure based on SURF.

## 2. Speeded-Up Robust Features (SURF)

SURF outperforms or is comparable to existing schemes in terms of repeatability, distinctiveness, and robustness, with much faster performance. This algorithm describes the keypoint detector and descriptor. The detector locates the keypoints in the image, and the descriptor describes the features of the keypoints and constructs the feature vectors of the keypoints.

### 2.1 Keypoint detector

SURF uses the determinant of the approximate Hessian matrix as the base of the detector. Integral images are used in Hessian matrix approximation, which allows fast evaluation of box filters. The integral image representation  $J$  of an image  $I$  is defined as:

$$J(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (1)$$

Given a point  $X = (x, y)$  in an image  $I$ , the Hessian matrix  $H(X, \sigma)$  in  $X$  at scale  $\sigma$  is defined as:

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (2)$$

where  $L_{xx}(X, \sigma)$  refers to the convolution of the second order Gaussian derivative  $\frac{\partial^2 g(\sigma)}{\partial x^2}$  with the image at point  $X = (x, y)$  and similarly for  $L_{yy}$  and  $L_{xy}$ . These derivatives are known as the Laplacian of Gaussians.

To reduce the computation time, a set of  $9 \times 9$  box filters (Fig. 2) is used as the approximations of a Gaussian with  $\sigma = 1.2$  and represents the lowest scale for computing the blob response maps. We will denote them by  $D_{xx}$ ,  $D_{yy}$  and  $D_{xy}$ . The weights applied to the rectangular are kept simple for computational efficiency. This yields:

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (3)$$

where  $w$  is a weight for the energy conservation between the Gaussian kernels and the approximated Gaussian kernels, and

$$w = \frac{|L_{xy}(1.2)|_F |D_{yy}(9)|_F}{|L_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.912\dots \approx 0.9 \quad (4)$$

$|x|_F$  is the Frobenius norm.

In order to detect keypoints using the determinant of Hessian it is necessary to introduce the notion of a scale space. The SURF constructs a pyramid scale space, like the SIFT. Different from the SIFT to repeatedly smooth the image with a Gaussian and then sub-sample the image, the SURF directly changes the scale of box filters to implement the scale space due to the use of the box filter and integral image.

## 2.2 Keypoint descriptor

The SURF used the sum of the Haar wavelet responses to describe the feature of a keypoint. Haar wavelets are used for the integral images to increase robustness and decrease computation time. Fig. 3 shows the Haar wavelets used to compute the responses in the  $x$  and  $y$  directions.

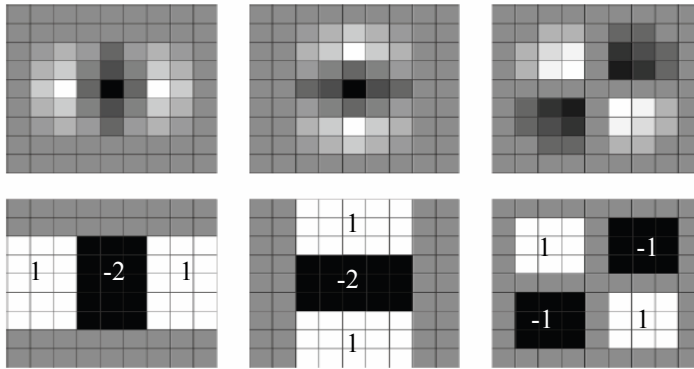


Fig. 2 Top row (left to right): The Gaussian second order partial derivative in  $L_{xx}$ ,  $L_{yy}$ , and  $L_{xy}$ . Bottom row (left to right): The approximation for the second order Gaussian partial derivative in  $D_{xx}$ ,  $D_{yy}$ , and  $D_{xy}$ .



Fig. 3 Left and right filters for computing the response in the  $x$  and  $y$  directions, respectively. Weights are 1 for black regions and -1 for white regions.

For the extraction of the descriptor, the first step consists of constructing a square region centered at the keypoint and oriented along the orientation decided by the orientation selection method introduced in [17]. The region is split up equally into smaller  $4 \times 4$  square sub-regions. For each sub-region, the Haar wavelet responses are computed at  $5 \times 5$  regularly spaced sample points (as shown in Fig. 4). We call  $d_x$  the Haar wavelet response in horizontal direction and  $d_y$  the Haar wavelet response in vertical direction. The keypoint descriptors are shown in Fig. 5.

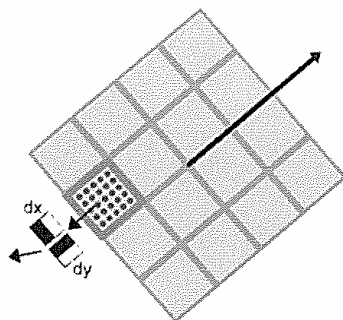


Fig. 4 The presentation of descriptor building.

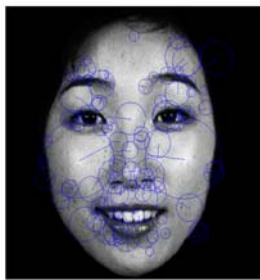


Fig. 5 Keypoint descriptors obtained using SURF.

Then, the wavelet responses  $d_x$  and  $d_y$  are summed up over each sub-region and form a first set of entries in the feature vector. In order to bring in information about the polarity of the intensity changes, we also extract the sum of the absolute values of the responses,  $|d_x|$  and  $|d_y|$ . Hence, each sub-region has a four-dimensional descriptor vector  $v$  for its underlying intensity structure  $v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$ . Each sub-region contributes four values to the descriptor vector leading to an overall vector of length  $4 \times 4 \times 4 = 64$ .

### 3. Feature extraction

#### 3.1 Feature vector normalization

A keypoint descriptor represents the gradient magnitude and orientation at each keypoint in a region around a keypoint location. Each keypoint descriptor (i.e., feature vector) is normalized to the unit length and a PDF descriptor (Fig. 6) is generated.

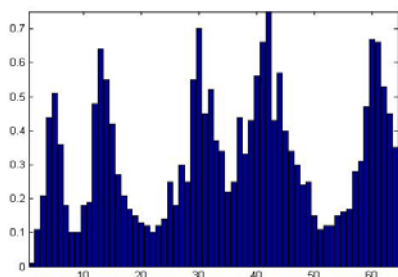


Fig. 6 PDF descriptor.

#### 3.2 KL divergence matching

Recently, some keypoints based matching methods generated impressive results in object recognition [24]. The matching of images in order to establish a measure of their similarity is a key problem in facial expression recognition. The measure of similarity we can associate to the features extracted using SURF directly result from the choice of independence between them. Independence of a set of random variables is statistically defined as the equality of the joint distribution of the variables and the product of their marginal PDFs. KL

divergence is thus used to compare the actual distributions and those of the statistics model [20]. KL divergence (5) has been employed to calculate the distance between the two PDF descriptors.

$$k(y_1, y_2) = \frac{1}{n} \sum_{i=1}^n y_1(x_i) \log\left(\frac{y_1(x_i)}{y_2(x_i)}\right) \quad (5)$$

where  $n$  is the number of bins and  $x_i$  is the  $i$ th bin.

The minimum distance between the two PDF descriptors is found using:

$$D_j = \arg \min_{y_j} \{k(y_i, y_j)\}, \quad y_j \in \{y_1, y_2, \dots, y_m\} \quad (6)$$

where  $y_i$  is a descriptor in the  $i$ th image,  $y_j$  is one of the  $m$  descriptors in the second image, and  $D_j$  is a descriptor in the  $j$ th image.

### 3.3 Recognition tally

Some suitable features were selected for increasing the distinguishing efficiency of the PDF descriptors. The recognition among the classes in a space can be measured using class separability. The PDF descriptors are utilized as a class separability measure. The recognition tally for each descriptor  $y_i$  from class  $A$  is defined as:

$$T(y_i) = \frac{[\sum_{B=1}^{C-1} (N_B \sum_{j=1}^{N_B} k(y_i, D_j))]/(C-1)}{(N_A \sum_{j=1}^{N_A} k(y_i, D_j))}, \quad A \neq B \quad (7)$$

where  $C$  is the number of classes and class  $B$  has  $N_B$  samples.

The denominator represents the minimum PDF descriptor of class  $A$ , such as happy. The numerator represents the minimum PDF descriptor of the other classes, such as angry, etc.

### 3.4 PDF descriptor extraction

Each training image is divided into a  $4 \times 4$  uniform grid. According to the recognition tally, the four largest PDF descriptors of each grid are extracted from the training image of each class (see Section 6.2 for details). These PDF descriptors are treated as recognition patterns of the test image. Different classes respond to different PDF descriptors.

## 4. Classification

The  $4 \times 4$  uniform grids for each image and the four largest PDF descriptors are chosen

for each expression class (Fig. 7). These PDF descriptors are used to discriminate facial expression  $F$  for each grid. An example is shown in below.

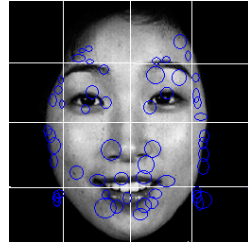


Fig. 7 Example of 4×4 uniform grids and four largest PDF descriptors.

$$F = \begin{array}{|c|c|c|c|} \hline \text{An.} & \text{Di.} & \text{Sa.} & \text{Ha.} \\ \hline \text{Di.} & \text{Sa.} & \text{Sa.} & \text{Sa.} \\ \hline \text{An.} & \text{Fe.} & \text{Sa.} & \text{Sa.} \\ \hline \text{Su.} & \text{Fe.} & \text{Sa.} & \text{Fe.} \\ \hline \end{array}$$

The central part of a face contains most of the important information. A Gaussian mask  $G$  on the grids is selected and the central part of the grids is assigned a heavier weight.

$$G = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 2 & 1 \\ \hline 2 & 4 & 4 & 2 \\ \hline 2 & 4 & 4 & 2 \\ \hline 1 & 2 & 2 & 1 \\ \hline \end{array}$$

Classification depends on the cumulative matching scores of each grid. The total scores of each facial expression for all grids are calculated to determine the highest score of facial expression in each image. An example of WMV classifier is shown in Fig. 8.

$(F)$				$(G)$						
An.	Di.	Sa.	Ha.	1	2	2	1	× =>	Ha.	1
Di.	Sa.	Sa.	Sa.	2	4	4	2		Fe.	7
An.	Fe.	Sa.	Sa.	2	4	4	2		Su.	1
Su.	Fe.	Sa.	Fe.	1	2	2	1		Sa.	20
								Di.	4	
								An.	3	

Fig.8 Example of WMV classifier.

## 5. Results

### 5.1 Data set

The JAFFE [25] database was used to evaluate the facial expression recognition performance of the proposed method. In the JAFFE database, there are 10 persons (subjects). For each subject, there are six types of facial expression: angry (AN), disgust (DI), fear (FE), happy (HA), sadness (SA), and surprise (SU). There are three or four samples corresponding to each facial expression of each person.

During training and testing, the face region was manually cut and resized to  $256 \times 256$  pixels according to the distance between two eyeballs (Fig. 9).

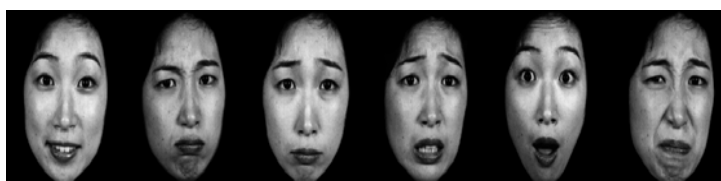


Fig. 9 Preprocessing of images (left to right: happy, angry, sadness, fear, surprise, disgust).

### 5.2 Simulation results

To demonstrate the efficiency of the proposed method, extensive experiments were conducted on the JAFFE database. Firstly, three samples of each facial expression per subject were selected. The training images were selected randomly and the rest were used for testing. Secondly, all the data in one subject were used as the test data, and the remaining subjects were used as training samples. The experiments were repeated ten times, each time using a different subject as the test data. The confusion matrixes of the average recognition rate are presented in Tables 1 and 2.

Table 1 Recognition rates of the proposed method (same subject)

Input/ Output	Angry (%)	Happy (%)	Surprise (%)	Sad (%)	Disgust (%)	Fear (%)
Angry	100	0	0	0	0	0
Happy	0	100	0	0	0	0
Surprise	0	0	100	0	0	0
Sad	0	0	0	90	0	10
Disgust	0	0	0	0	90	10
Fear	10	0	0	0	0	90



Table 2 Recognition rates of the proposed method (different subject)

Input/ Output	Angry (%)	Happy (%)	Surprise (%)	Sad (%)	Disgust (%)	Fear (%)
Angry	100	0	0	0	0	0
Happy	0	100	0	0	0	0
Surprise	0	0	100	0	0	0
Sad	0	0	0	90	10	0
Disgust	0	0	10	0	80	10
Fear	0	0	0	10	0	90

### 5.3 Results comparison

SIFT [11-14] is one of the efficient detecting and matching features of state-of-the-art method. The implementation of SURF was replaced by SIFT and the difference in performance was determined. A given training and testing set was used for both methods. The algorithms were implemented in MATLAB and run on a PC with an Intel Core 2 Duo 2.4-GHz CPU and 2 GB of RAM.

Two accuracies for same and different subjects, respectively, for the two methods are shown in Fig. 10 and Fig. 11. The average rate is compared in Table 3. The average time of the training and the testing is compared in Table 4. The experiment results show that the proposed method recognition rate is better than that of SIFT

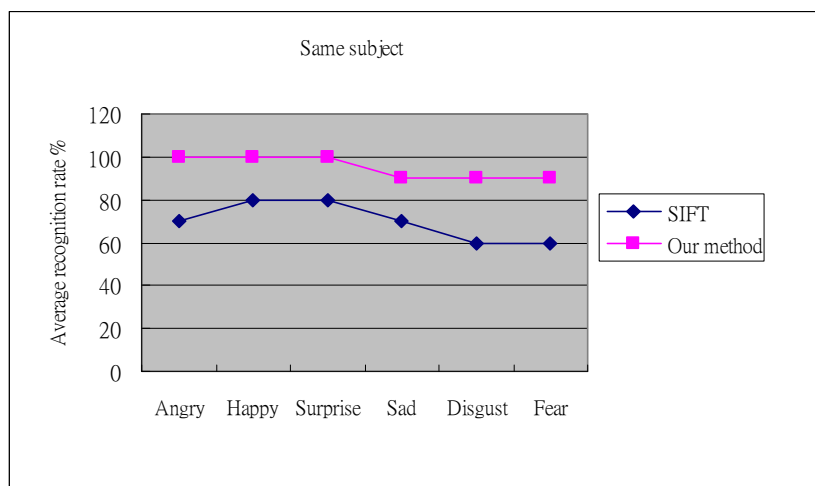


Fig. 10 Comparison of recognition rates for two algorithms for same subjects.

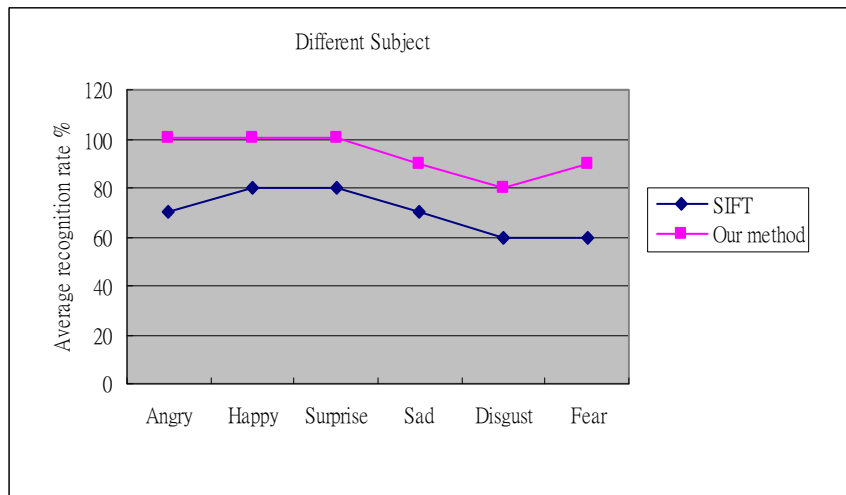


Fig. 11 Comparison of recognition rates for two algorithms for different subjects.

Table 3 Average recognition rate

		Same subject		Different subject	
		SIFT	Our	SIFT	Our
Recognition Feature	Angry	70	100	70	100
	Happy	80	100	80	100
	Surprise	80	100	80	100
	Sad	70	90	70	90
	Disgust	60	90	60	80
	Fear	60	90	60	90
Average (%)		71.67	95	71.67	93.33

Table 4 Average time of facial expression recognition (second)

	SIFT	Our Method
Training (per subject)	262.393	19.897
Testing (per image)	0.828	0.63

#### 5.4 Performance comparison with other methods

Table 5 compares the result obtained using the proposed methods with those obtained using existing methods for the given database. It is noted that the training and test datasets are different. The proposed approach achieves the best performance.

Table 5 Comparison with existing methods

Reference	Method	Accuracy (%)
Shinohara [9]	HLAC + Fisher weight maps	69.4
Lyons [26]	Wavelet + PCA + LDA	75
Huang, M.W. [27]	GPLVM + SVM	65.24
Mingwei Huang [28]	SNE + SVM	73
Bin Hua [29]	bidirectional 2DPCA	92.52
Our Method (same subject)	SURF + KL divergence + WMV	95
Our Method (different subject)	SURF + KL divergence + WMV	93.33

## 6. Conclusion and Discussion

### 6.1 Conclusion

A facial expression recognition method was proposed. It detects and computes keypoint descriptors (feature vectors) of an image using SURF. The keypoint descriptor is normalized to a PDF descriptor. Then, KL divergence is used to calculate the distance between two PDF descriptors to determine the two nearest PDF descriptors from two images. The recognition tally is calculated from the PDF descriptors and used for initial classification. Then, WMV is used for final classification. Experimental results show excellent performance in recognizing facial expressions.

### 6.2 Discussion

The four largest PDF descriptors for each grid are selected because the accuracy of the recognition rate depends on the number of PDF descriptors. However, an excessive number of PDF descriptors degrades recognition performance. The method used for determining the optimal number of PDF descriptors is illustrated in Fig. 12. The figure is generated from each selection of PDF descriptors from the JAFFE database using the SURF method. The recognition rate is highest when the four largest PDF descriptors are selected.

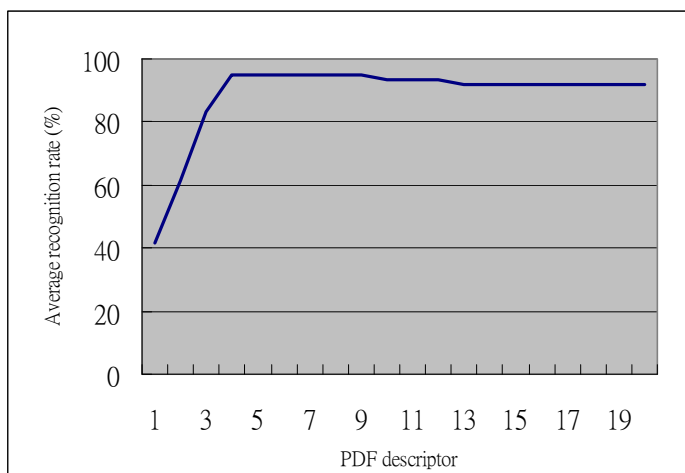


Fig. 12 Extraction of PDF descriptors information.

## Acknowledgment

The authors would like to thank the JAFFE Technical Agent. We are also grateful to the anonymous reviewers and editors for their valuable comments.

## References

- [1] A. Mehrabian, "Communication without words", *Psychol. Today*, vol. 2, no. 4, pp. 53–56, 1968.
- [2] M.Pantic and L.Rothkrantz, "Facial action recognition for facial expression analysis from static face images", *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 3, pp. 1449–1461, Jun. 2004.
- [3] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu, "A principal component analysis of facial expression", *Vis. Res.*, vol. 41, no. 9, pp. 1179–1208, Apr. 2001.
- [4] Fukunaga K. (1990) *Introduction to Statistical Pattern Recognition*, 2nd edn. Academic Press, Boston.
- [5] B. Raducanu, J. Vitria, "Online nonparametric discriminant analysis for incremental subspace learning and recognition", *Pattern Anal Applic*, vol. 11, pp. 259-268, 2008.
- [6] S.C. Tai and H.F. Huang, "Facial expression recognition in video sequences", *Sixth Int'l Symposium on Neural Network*, pp. 1026-1033, Wuhan, China, May 2009.
- [7] S.C. Tai, H.F. Huang and K. C. Chung, "Automatic Facial Expression Discrimination System", *Far East Journal of Electronics and Communications*, vol. 1, issue1, pp. 23-31, 2007.
- [8] S. Negahdaripour, "Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 961-979, 1998.

- [9] Y. Shinohara and N. Otsu, "Facial Expression Recognition Using Fisher Weight Maps", *IEEE Conf. on Automatic Face and Gesture Recognition*, pp. 499-504, 2004
- [10] T. O., M. P. and T. M., "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, No. 7, pp. 971-987, 2002.
- [11] M. Wen, Y. Li, Q. Zhuo and W. Wang, "Novel scale-invariant keypoint detector", *Opt. Eng.*, vol. 47(5), pp. 057203-1- 057203-6, May 2008.
- [12] J. Luo, Y. Ma, E. Takikawa, S. H. Lao, M. Kawade, B. L. Lu, "Person-Specific SIFT features for face recognition", *International Conference on Acoustic, Speech and Signal Processing*, 2, pp. 593-596, Hawaii, 2007.
- [13] Z. Li, J-I IMAI, and M. KANEKO, "Facial Feature Localization Using Statistical Models and SIFT Descriptors", *The 18th IEEE Int Symposium on Robot and Human Interactive Communication*, pp. 961-966, Japan, 2009.
- [14] H. Soyel and H. Demirel, "Facial expression recognition based on discriminative scale invariant feature transform", *ELECTRONICS LETTERS*, vol. 46, No. 5, 4th March 2010.
- [15] Geng Du, Fei Su, Anni Cai, "Face recognition using SURF features", *MIPPR 2009: Pattern Recognition and Computer Vision*, vol. 7496, pp. 749628.1- 749628.7, 2009.
- [16] D. Sargent, C.I. Chen, C.M. Tsai, Y.F. Wang and D. Koppel, "Feature Detector and Descriptor for Medical Images", *Proc. SPIE*, vol. 7259, pp. 7259Z.1-7259Z.8, 2009.
- [17] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "Speeded-up robust features (SURF)", *Comput. Vis. Image Underst.*, 110(3), 346-359, 2008.
- [18] T. Terriberry, L. French and J. Helmsen, "GPU Accelerating Speeded-Up Robust Features", *3D Data Processing, Visualization and Transmission*, Georgia Tech, Atlanta, June 2008.
- [19] Solomon Kullback, *Information Theory and Statistics*, Dover Publications Inc., Mineola, New York, 1968.
- [20] Z. Wenchao, S. Shiguang, C. Xilin and G. Wen, "Local gabor binary patterns based on kullback-leibler divergence for partially occluded face recognition", *IEEE Signal Processing Letters*, vol. 14, no. 14, pp. 875-878, 2007.
- [21] Nick Littlestone and Manfred K. Warmuth, "The weighted majority algorithm", *30th Annual Symposium on Foundations of Computer Science*, pp. 256-261, Oct. 1989.
- [22] E. Bauer, R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants", *Machine Learning*, vol. 36, issue. 1-2, pp. 105-142, 1999.
- [23] X. Artaechevharria, A. Munoz-Barrutia and C. Ortiz-de-Sollrzano, "Efficient classifier generation and weighted voting for atlas-based segmentation: two small steps faster and closer to the combination oracle", *Proc. SPIE*, vol. 6914, pp. 69141W.1-69141W.9, 2009.
- [24] K. Mikolajzyk and C. Schmi, "A performance evaluation of local descriptors", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1615-1630, 2005.
- [25] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba, "Coding Facial Expressions with Gabor Wavelets", *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, Apr. 14-16, 1998.

- [26] M.Lyons, J.Budynek, and S.Akamastu, "Automatic Classification of Single Facial Images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 1357-1362, 1999.
- [27] Huang, M.W.; Wang, Z.W.; Ying, Z.L., "A novel method of facial expression recognition based on GPLVM Plus SVM", *ICSP*, pp. 916-919, 2010.
- [28] Mingwei Huang; Zhen Wang; Zilu Ying, "Facial expression recognition using Stochastic Neighbor Embedding and SVMs", *ICSSE*, pp. 671-674, 2011.
- [29] Bin Hua; Ting Liu, "Facial Expression Recognition Based on Local Feature Bidirectional 2DPCA", *ITCS*, vol. 1, pp. 301-304, 2009.