Pamukkale Univ Muh Bilim Derg, 23(5), 527-531, 2017

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

Facial expression recognition based on compressive sensing and pyramid processing

Sıkıştırılmış algılama ve piramit işlemeye dayalı yüz ifade tanıma

Alaa ELEYAN[1], Abubakar M. ASHIR[2]*

[1]Electrical & Electronics Engineering Department, Engineering & Architecture Faculty, Avrasya University, Trabzon, Turkey.
aeleyan@avrasya.edu.tr
[2]Electrical & Electronics Engineering Department, Engineering Faculty, Selçuk University, Konya, Turkey.
ashir4real@yahoo.com

Received/Geliş Tarihi: 04.04.2016, Accepted/Kabul Tarihi: 01.09.2016 doi: 10.5505/pajes.2016.43402
* Corresponding author/Yazışılan Yazar
Research Article/Araştırma Makalesi

Abstract

In this paper, a new approach has been proposed for improved facial expression recognition. The new approach is inspired by the compressive sensing theory and multi-resolution approach to facial expression problems. Initially, each image sample is decomposed into desired levels of its pyramids at different sizes and resolutions. At each level of the pyramid, features are extracted using a measurement matrix based on compressive sensing theory. These measurements are concatenated together to form a feature vector for the original image. The results obtained from the approach using three distance measurement classifiers (Manhattan, Euclidean, Cosine) and support vector machine are impressive and outperforms most of its counterpart algorithms in the literature using the same databases and settings.

Keywords: Facial Expression recognition, Compressive sensing, Image pyramid

Öz

Bu makalede, geliştirilmiş yüz ifadesi tanıma için yeni bir yaklaşım önerilmiştir. Bu yeni yaklaşım sıkıştırma algılama teorisinden ve yüz ifadesi problemine çoklu çözünürlük yaklaşımından esinlenmektedir. Başlangıçta, her bir görüntü örneği farklı boyutlarda ve çözünürlüklerdeki piramitlerin istenilen seviyesine ayrıştırılmaktadır. Piramidin her seviyesinde, özellikler sıkıştırma algılama teorisine dayanan bir ölçüm matrisi kullanılarak ayrıştırılmaktadır. Bu ölçümlerin tamamı orijinal görüntü için bir özellik vektörü oluşturmak için bir araya getirilmektedir. Üç uzaklık ölçümü sınıflandırıcısı (Manhattan, Öklid, kosinüs) ve destek vektör makinesi kullanımından elde edilen sonuçlar, aynı veri tabanları ve ayarlarının kullanıldığı literatürdeki benzer algoritmaların çoğundan daha etkileyici ve iyidir.

Anahtar kelimeler: Yüz İfade tanıma, Sıkıştırma algılama, Görüntü piramidi

1 Introduction

Facial Expression Recognition (FER) is one of the branches of Pattern Recognition (PR) which enjoys increasing patronage from many works of life in recent times. This could be attributed to the developments in technology and human's needs for information and intelligence gathering. Some of the emerging applications of FER are in marketing, security, psychology, medical diagnosis, human-machine interaction and entertainments [1]. The algorithm flow for FER is not much different from its counterpart algorithms in PR. The steps include: pre-processing, feature extraction, classification and the decision. Generally, in FER two major approaches are adopted for feature extraction. First, is the Component-based (holistic) and the second is feature-based (local) approach. In the former, the entire face image is used as input to extract features while in the later only some key points within the face image (e.g. eye, nose, mouth e.tc) are used to take some geometrical measurements and localized information around them [2],[3].

Due to the tasking nature of FER, complex techniques have been used by researchers to obtain more robust and distinct features for the various emotions recognition. Use of multi-resolution algorithms such as Gabor Wavelets Transform (GWT), Discrete Wavelets transform (DWT) to mention but few, are very common in FER and appears to have an edge over other feature extractors like Local Binary Pattern (LBP), Principal Component Analysis (PCA) and Local Discriminant Analysis (LDA) [2]. Authors in [2], used a multi-resolution transform called Curvelets Transform (CT) at different orientations and scales to form Curvelets products which were wrapped around their origin. The products were then used to extract Curvelets coefficients using inverse CT. The coefficients are subsequently used as feature vectors. Though improved performance has been reported but intensive computations are required to arrive at that performance. In a similar way, in [3]-[8] authors used GWT in one form or the other to encode features for FER. For instance, [3] subjected the face images to local, multi-scale Gabor-filter operations, and then the resulting Gabor decompositions were encoded using radial grids, imitating the topographical map-structure of the Human Visual Cortex (HVC). Due to the similarity of Gabor filter response to HVC and its performance, many local variants of the algorithms were used recently [4],[7],[8]. In general, these algorithms come with additional cost of computation, extensive memory usage and in most cases feature vector dimensionality reduction becomes a necessity.

In this paper, feature extraction algorithm based on Compressive Sensing (CS) theory and image pyramid processing has being proposed. Initially, image pyramids were computed at different layers with Gaussian approximations filter. The pyramid layers were concatenated. Final features are extracted from the concatenated vectors using CS theory. The approach has the combined multi-resolution capability, simplicity and higher performance.

The rest of the paper is organized as follows: Section II briefly discusses image pyramid processing and compressive sensing theory. Section III outlines the proposed approach and section

IV presents the experimental results. Section V concluded the findings.

2 Feature extraction

Feature extraction is a crucial stage in any pattern recognition problem. The technique deployed to extract features from the original image samples are briefly explained.

2.1 Image pyramid

One of the powerful but yet conceptually simple structure for representing image in more than one resolution is image pyramid. The motivation for image pyramid and of course, any other multi-resolution algorithm is to view salient features of image at different resolutions so that, one feature which is difficult to detect in one resolution might easily be found in another resolution [9]. For instance, small or low contrast objects reveal more details at higher resolutions whereas for large or high contrast objects within an image, a coarse resolution (lower) is all that is needed [9]. In a situation where both low and high contrast objects coexist (which is usually the case), viewing the image at different resolutions will enhance the chance for more features revelation.

An image pyramid is a collection of decreasing resolution images arranged in pyramidal shape. As you move up from the base of the image pyramid both resolution and size of the image decreases. A succeeding level of the pyramid is obtained from its preceding level by applying an approximation filter (low-pass filter) and down sampling the results both along rows and columns. For applications like image compression, original image can be reconstructed from its pyramids using level residual and interpolation filters [9]. Figure 1 and 2 showed schematic of image pyramid decomposition and an example of samples image being decomposed into four level pyramids.

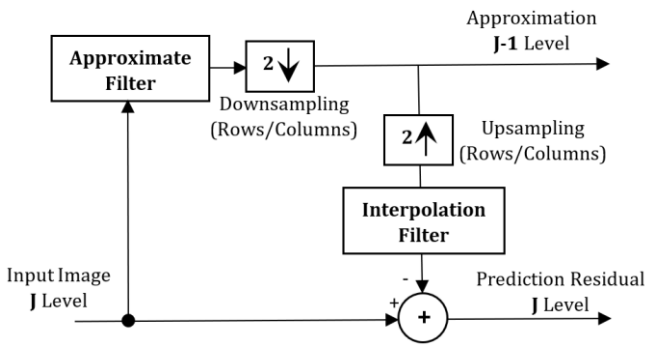

Figure 1: Pyramid decomposition schematic.

Figure 2: A sample decomposed into 4-level pyramids using Gaussian approximation filter.

In the pyramid decomposition stage, a low pass two dimensional Gaussian filter can be used. The Gaussian kernel's weights can be constructed by using (1). The mean distribution of the filter is centered at the middle of the kernel i.e. $k(x,y) = k(0,0)$, where x and y are the kernel's horizontal and vertical coordinates respectively.

$$k(x,y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}+\frac{(y-\mu)^2}{2\sigma^2}\right]} \quad (1)$$

2.2 Compressive sensing

The Nyquist-Shannon sampling theorem [10] is considered as one of the cornerstones theorems in communication and digital signal processing world. It provided hints on how a continuous-time signal can be flawlessly recovered from its original samples. The theory states that a continuous-time signal can be reconstructed or recovered from its regularly space samples when the sampling frequency is two times larger or equal to its bandwidth [10]. Sampling at Nyquist rate is easy to implement however, the question of efficiency becomes of great concern in terms of data rates to be collected. Sampling at Nyquist rate usually leads to collection of large amount of data samples [11]. Related work by Shannon [12] in information theory used the idea of entropy to prove that the sampled data could be compressed to fewer samples since substantial parts of it are redundant data and does not contain the needed information. Having realized that, number of data compression algorithms evolved over times to trim down data redundancy. For example, in JPEG compression, the signal which is the digitized *2-dimensional* image sampled from the *2-dimensional* image sensors grid, is transformed into the Discrete Cosine Transform (DCT) domain. In the DCT transform domain, most of the samples with small or negligible coefficients (i.e. scales or amplitude) are thrown away [11],[13],[14].

Compressive sensing (CS) is one of the data compression algorithms with even more radical approach to data compression [13]. In the transformation domain a signal of length N is represented in terms of its scaling vector and its basis functions. The signal in this domain is K-Sparse if it has maximum K nonzero and $(N-K)$ zero coefficients in its scaling vector. The crucial point in CS theory is that when K is way much less than N, i.e. the signal is compressible in the transform domain [11],[13],[14].

In CS, the aim is to reduce number of regularly spaced samples taken from the original signal. This is achieved by taking values of the signal at certain places instead of all the samples. The new samples are referred to compressed measurements y and they are calculated as follows:

$$y = \phi x = \phi.\psi.s = \theta.s \quad (2)$$

Where ϕ is the $M \times N$ measurement matrix, $M \ll N$, and s is the scaling coefficients of the signal x in the transform domain, ψ. The original signal x reconstruction from its compressed measurements y is achieved through solving optimization problem and not simply by inverse transformation techniques [11].

An important property of the measurement matrix ϕ, is that it does not need to have a specific structure like transformation matrices or sampling matrices. In fact, authors in [11], stated that the measurement matrix only needs to satisfy the Restricted Isometric Property (RIP) for a given number of measurements. They also proved that a random matrix with independent and identically distributed (i.i.d.) Gaussian random variables as its elements will satisfy the RIP property [11].

Recovering the original signal from these compressed measurements is another topic. However, within the scope of this paper, the relevant part is the decomposition or the sampling part of the CS framework. Since the compressed measurements are guaranteed to have preserved the salient information from the original signal, it can be used as features. Therefore, reconstruction is not discussed here.

3 Proposed approach

The proposed approach blended both the simplicity and strength of multi-resolution algorithms and data compression capabilities (dimensionality reduction) of CS. The general flow of the algorithm is shown in Table 1.

Table 1: Proposed approach.

1.	Given a training set of M images: set a matrix $\mathbf{T} = [\mathrm{I}_1, \mathrm{I}_2, \dots, \mathrm{I}_M]$; such that I_i is a column version of the ith image,
2.	For each image in $\mathbf{T}$ generate its pyramid to a predefined J levels using Gaussian approximation filter,
3.	For each level of image pyramid perform the following:
4.	Normalize images $\mathrm{I}_{1,i}, \dots \mathrm{I}_{J,i}$ within the range [0 1],
5.	Concatenate image pyramid, $\mathrm{I}_{1,i}, \dots \mathrm{I}_{J,i}$, corresponding to I_i image to form new column representation Γ_i.
6.	Form the matrix $\mathbf{R} = [\Gamma_1, \Gamma_2, \dots, \Gamma_M]$,
7.	Compute vector $\Omega = [\mu_1, \mu_2, \dots \mu_M]$. Where μ_i represents the arithmetic mean of pyramidal image Γ_i.
8.	Form a Difference matrix $\mathbf{A} = [\Gamma_1 - \mu_1, \Gamma_2 - \mu_2, \dots, \Gamma_M - \mu_M]$,
9.	Project matrix $\mathbf{A}$ into Gaussian Space by using (2),
10.	Return the feature vectors,
11.	Classify the data,
12.	Decision,
13.	end.

4 Simulation results

Experiments were carried out on two different facial expression databases and results were tabulated using the new proposed approach. Two standard methods of training and cross-validation of the results were adhered to. *Leave-One-Pose-Out* (LOPO) cross validation was used on Japanese Female Facial Expression (JAFFE) database [5] and *n-fold* cross validation on Cohn-Kanade (CK) database [15]. Both methods are frequently used in literatures [1]. The experimental constant parameters used are listed below:

- Filter size: 9×9,
- Gaussian Filter mean: $\mu = 0$,
- Gaussian Filter Standard Deviation: $\sigma = 20$.

4.1 Leave-One-Pose-Out cross-validation on JAFFE.

The Jaffe database has 213 images from 10 subjects each having 3 to 4 sample images per expression. 210 images were used in this context. Before the training all the samples were grouped into 7 classes based on expression content (i.e. person-independent training). One sample from each expression's group was used as test data while the remaining samples were used as training set. The process was repeated with different sample, until each sample was uniquely used as testing data as well as training data. The recognition rate was given as the average of all the runs. Figure 3 contains a cross section of the JAFFE database. The vertical columns represent person-independent expression classes whereas the rows represent person-dependent expression classes for a particular subject.

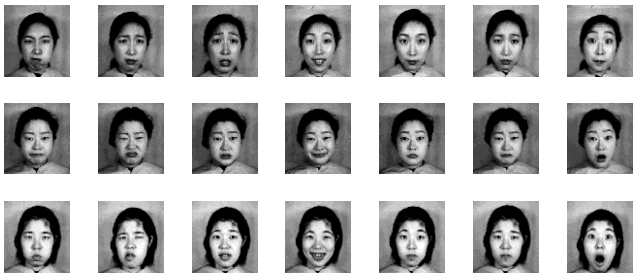

Figure 3: JAFFE database sample images.

4.2 *n-fold* cross validation on CK.

LOPO cross validation technique could be very exhaustive and computationally intensive [1] in a database with many samples like CK. Since CK database is very huge containing 97 subjects and total of 8795 sample images [16], 10-fold cross validation method of training was used. In 10-fold cross validation, all the images were initially grouped into seven classes based on expression information (person-independent), then equal number of samples were randomly drawn (without duplication) from each expression class to form 10 equally sized groups or folds. During the training, one fold was used as the testing set whereas the remaining nine folds were used as training set. The procedure was repeated 10 times, each time with a different fold. The recognition rate was given as the average of the 10 runs conducted.

It is worth noting that original CK database contains 640×490 sized samples with non-facial background information. In this experiment, all face samples were cropped to remove irrelevant background information and resized to 256×256. Since every expression in CK database starts from neutral to peak level, neural samples were not included in an expression that is not neutral. Examples of preprocessed images from CK database are shown in Figure 4.

Figure 4: CK database sample images.

Table 2 and 3 contained the simulation results of the proposed approach using LOPO on JAFFE database. The results were obtained using level 1 and 2 of the pyramid at different feature vector length N, and four different classifiers namely: Manhattan ($\boldsymbol{\delta}_{l_1}$), Euclidean ($\boldsymbol{\delta}_{l_2}$), Cosine ($\boldsymbol{\delta}_{cos}$) and support vector machine (SVM). Whereas Table 4 and 5 present similar results on CK database using *10-fold* cross validation. Results from other approaches are compared in Table 6 and 7 with the results obtained in this work under the same scenario.

4.3 Discussion

The combination of CS and pyramids processing has achieved a better performance compared to its counterparts with great simplicity. In Table 2 to 5 for example, in both JAFFE and CK

database, increase in pyramid layer to a certain level improves the performance. However, for distance measurement classifiers the improvement is not much and sometimes increase in the pyramid layers beyond certain number does not necessarily translate into improvement in performance. In fact, experiments have shown that 2 layers of the pyramid are adequate and any attempt to go beyond that, a slight degradation in performance was observed. This may be due to that fact that, for the sample images used no discernible information could be extracted beyond the second layer of the pyramid. As a result use of many pyramids layers for those may be more of a dimensionality curse than of an advantage. Moreover, while this is true for both classifiers, a distinct response to the pyramid layer increase was exhibited by SVM which is different from the rest of the distance measurement classifiers. Distance measurement based classifiers results improve slightly when both the feature vector length and pyramid layer are increased. It might be connected to the fact that they do not require any learning to be able to adequately encode any additional information which might be available at lower resolutions. In sharp contrast to distance measurement classifier, SVM classifier both responds rapidly to the increase in pyramid layers and feature vector length. For instance, in JAFFE database higher performances are recorded up to 700 feature vector lengths after which the performance begins to decline. Whereas for CK database, lower feature vector length appears to be more suitable.

Table 2: LOPO recognition rate (%) on JAFFE with 1 layer.

N	δ_{l_1}	δ_{l_2}	δ_{cos}	SVM
50	84.76	83.33	81.90	45.71
100	87.14	87.14	84.76	61.43
300	88.57	89.05	88.57	75.24
500	87.62	88.10	87.62	82.86
700	89.57	89.54	89.05	83.33
900	86.19	86.67	86.67	80.95

Table 3: LOPO recognition rate (%) on JAFFE with 2 layers.

N	δ_{l_1}	δ_{l_2}	δ_{cos}	SVM
50	85.24	84.76	87.14	64.76
100	86.67	87.14	88.57	79.05
300	86.67	87.14	88.57	87.14
500	88.57	88.57	89.56	85.71
700	87.62	87.14	89.52	90.48
900	87.14	86.67	88.10	88.57

Table 4: 10-fold recognition rate (%) on CK with 1 layer.

N	δ_{l_1}	δ_{l_2}	δ_{cos}	SVM
50	98.30	98.30	98.30	88.20
100	98.28	98.28	98.28	88.20
200	98.36	98.36	98.36	55.66
300	99.40	98.40	99.40	39.67
400	99.36	99.36	99.36	28.89
500	99.32	99.32	99.32	21.09

Table 5: 10-fold recognition rate (%) on CK with 2 layers.

N	δ_{l_1}	δ_{l_2}	δ_{cos}	SVM
50	99.26	99.26	99.26	95.28
100	99.30	99.30	99.30	92.37
200	99.28	99.28	99.28	85.93
300	99.38	99.38	99.38	80.39
400	99.32	99.32	99.32	75.31
500	99.28	99.28	99.28	70.47

Table 6: Comparison with different approaches on JAFFE database.

Ref.	Feature Extractor	Classifier	R. Rate (%)
[3]	Radial Enc. Gabor	SVM	89.67
[16]	LPQ+es_LBP-s	SVM	76.67
[17]	LBP	SVM	53.80
[17]	LBP	SVM	65.71
[18]	Gabor Histogram	SVM	58.70
[19]	LBP	SVM	62.90
[20]	CT	SVM	63.81
[20]	ASM	Ensembles	68.50
[21]	LBP+Gabor	Fusion SVMs	70.00
[22]	Gabor Filter	Two level VMs	72.00
[23]	MB-LGBP	SVM	74.18
Proposed	CS+Pyramid	SVM	90.48

Table 7: Comparison with different approaches on CK database.

Ref.	Subjects	Classes	Validation	R. Rate (%)
[3]	94	7	10-fold	91.51
[24]	76	6	5-fold	90.9
[25]	90	6	–	93.66
[26]	97	6	10-fold	96.26
[27]	90	6	LOPO	96.33
[28]	96	7(6)	10-fold	88.4 (92.1)
[29]	90	7	10-fold	86.9
[30]	90	6	LOPO	93.8
[31]	97	6	–	93.8
Proposed	97	7	10-fold	99.3

4.4 Comparison

In Tables 6 and 7, comparisons of the validity of our proposed approach with others in the literature under the same settings were conducted. The best result obtained from the proposed approach is chosen to do the comparison.

5 Conclusion

A new approach for facial expression recognition has been developed. The proposed approach has attributes of simplicity and better performance with regards to recognition rate and computational complexity as compared to other approaches. The proposed approach edges out most of its counterparts in the literature which further proves its validity.

6 References

[1] Fasel I, Juergen L. "Automatic facial expression analysis". *The Journal of Pattern recognition Society,* 36(1), 259-275, 2003.

[2] Min T, Feng C. "Facial expression recognition and its application based on curvelet transform and PSO_SVM". *International Journal for Light and Electron Optics,* 124(22), 5401–5406, 2013.

[3] Wenfei G, Cheng X, Venkatesh YV, Dong H, Hai L. "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis". *The Journal of Pattern recognition Society,* 45(1), 80-91, 2012.

[4] Shiqing Z, Lemin L, Zhijin Z. "Facial expression recognition based on Gabor wavelets and sparse representation". *11th International Conference on Signal Processing,* Beijing, China, 21-25 October 2012.

[5] Michael J, Shigeru A, Miyuki K, Jiro G. "Coding facial expressions with Gabor wavelets". *3rd International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 14-16 April 1998.

[6] Shishir B, Ganesh K. "Recognition of facial expressions using gabor wavelets and learning vector quantization". *Journals of Engineering Applications of Artificial Intelligence,* 21(7), 1056–1064, 2008.

[7] Baochang Z, Shiguang S. "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition". *IEEE Transactions on Image Processing*, 16(1), 57–68, 2007.

[8] Yimo G, Zhengguang X. "Local Gabor phase difference pattern for face recognition". *19th International Conference on Pattern Recognition*, Tampa, FL, USA, 8-11 December 2008.

[9] Gonzalez RR, Woods RE. *Digital Image Processing.* 3rd ed. USA, Pearson publishers. 2008.

[10] Shannon C. "Communication in the presence of noise". *Proceedings of the Institute of Radio Engineers,* 37(1), 10-21, 1949.

[11] Emamnuel C. "Compressive sampling". *Proceedings of the International Congress of Mathematicians,* Madrid, Spain, 22-30 August, 2006.

[12] Shannon C. "A mathematical theory of communication". *Bell System Technical Journal,* 27(5), 379-423, 1948.

[13] Baraniuk R. "Compressed sensing [Lecture Notes]". *IEEE Signal Processing Magazine,* 124(2), 118-124, 2007.

[14] Eleyan A, Kose K, Cetin E. "Image feature extraction using compressive sensing". *Advances in Intelligent Systems and Computing*, 233, 177-184, 2013.

[15] Kanade T, Cohn J, Tian Y. "Comprehensive database for facial expression analysis". *4th International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, 28-30 March 2000.

[16] Wei-Lu C, Jian-Jiun D, Jun-Zuo L. "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection". *An International Journal of Signal Processing,* 117, 1-10, 2015.

[17] Ying Z, Fang X. "Combining LBP and adaboost for facial expression recognition". *9th International Conference on Signal Processing*, Beijing, China, 26-29 October 2008.

[18] Guo G, Dyer R. "Facial expression recognition based on Gabor histogram feature and MVBoost". *Journal of Computer Resource and Development,* 44(3), 1089-1096, 2007.

[19] Huang M, Wang Z, Ying Z. "A new Method for Facial Expression Recognition based on Sparse Representation plus LBP". *3rd International Congress on Image and Signal Processing,* Yantai, China, 16-18 October 2010.

[20] Cai L, Yin Z. "A new approach of facial expression recognition based on contourlet transform". *International Conference on Wavelet Analysis and Pattern Recognition,* Baoding, China, 12-15 July 2009.

[21] Zavaschi T, Koerich A, Oliveira L. "Facial expression recognition using ensemble of classifiers". *International Conference on Signal Acoustics, Speech and (ICASSP)*, Prague, Czech Republic, 22-27 May 2011.

[22] Shan C, Gong S. "Facial expression analysis across databases". *International Conference on Multimedia Technology*, Hangzhou, China, 26-28 July 2011.

[23] Zhang Z, Xu C, Wang J, Chen X. "Facial expression recognition based on MB-LGBP feature and multi-level classification". *Advances in Intelligent and Soft Computing,* 129, 37-42, 2011.

[24] Yeasin M, Bullot B, Sharma R. "From facial expression to level of interest: a spatio-temporal approach". *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 27 June-2 July 2004.

[25] Aleksic P, Katsaggelos A. "Automatic facial expression recognition using facial animation parameters and multi-stream HMMS". *IEEE Transactions on Information Forensics and Security,* 1(1) 3-11, 2006.

[26] Li Z, Imai J, Kaneko M. "Facial expression recognition using facial-component-based bag of words and phog descriptors". *The Journal of the Institute of Image Information and Television Engineers*, 64(2), 230-236, 2010.

[27] Shan C, Gong S, McOwan P. "Robust facial expression recognition using local binary patterns". *International Conference on Image Processing,* Genova, Italy, 14 September 2005.

[28] Zhao G, Pietik M. "Dynamic texture recognition using local binary patterns with an application to facial expressions". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 29(6), 915-928, 2007.

[29] Bartlett M, Littlewort G, Fasel I, Movellan R. "Real-time face detection and facial expression recognition: development and application to human computer interaction". *Conference on Computer Vision and Pattern Recognition Workshop,* Madison, Wisconsin, USA, 16-22 June 2003.

[30] Littlewort G, Bartlett M, Fasel I, Susskind J, Movellan J. "Dynamics of facial expression extracted automatically from video". *Journal of Image and Vision Computing,* 24(6), 615-625, 2006.

[31] Tian Y. "Evaluation of face resolution for expression analysis". *Conference on Computer Vision and Pattern Recognition Workshop,* Washington, DC, USA, 27 June-2 July 2004.