Enhancing ELM-based facial image classification by exploiting multiple facial views

Alexandros Iosifidis, Anastasios Tefas and Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
{tefas,pitas}@aiia.csd.auth.gr

Abstract
In this paper, we investigate the effectiveness of the Extreme Learning Machine (ELM) network in facial image classification. In order to enhance performance, we exploit knowledge related to the human face structure. We train a multi-view ELM network by employing automatically created facial regions of interest to this end. By jointly learning the network parameters and optimized network output combination weights, each facial region appropriately contributes to the final classification result. Experimental results on three publicly available databases show that the proposed approach outperforms facial image classification based on a single facial representation and on other facial region combination schemes.

Keywords: Extreme Learning Machine, Facial Image Classification, Multi-view Learning

1 Introduction

Extreme Learning Machine [7] is a relatively new algorithm for fast Single-hidden Layer Feedforward Neural (SLFN) network training that leads to fast network training requiring low human supervision. Conventional SLFN network training algorithms require the input weights and the hidden layer bias values to be adjusted using a parameter optimization approach, like gradient descend. However, gradient descend-based learning techniques are generally slow and may decrease the network's generalization ability, since they may lead to local minima. Unlike the popular thinking that the networks parameters need to be tuned, in ELM the input weights and the hidden layer bias values are randomly assigned. The network output weights are, subsequently, analytically calculated. Despite the fact that the determination of the network hidden layer output is a result of randomly assigned weights, it has been shown that SLFN networks trained by using the ELM algorithm have the properties of global approximators [5]. In addition, it has been recently shown that ELM networks are able to outperform other state-of-the-art classifiers, like Support Vector Machine (SVM). Due to its effectiveness and its fast learning process, the ELM network has been widely adopted in many classification problems, including facial image classification [4, 30, 6, 10, 13, 11, 16, 15].

Despite its success in many classification problems, the ELM training process by randomly assigning the network hidden layer weights and the corresponding bias values lacks the ability of learning optimized hidden network parameters from data. For example, a SLFN network trained on a facial image database by using the Backpropagation algorithm would be able to learn network hidden layer parameters highlighting the facial regions, e.g. the regions corresponding to the eyes or to the mouth, discriminating the persons in the database. On the other hand, the ELM algorithm can be considered as a learning process formed by two processing steps. The first step corresponds to a nonlinear mapping process of the input space to a (usually high-dimensional) feature space, noted as ELM space hereafter, preserving some properties of the training data. In the second step, an optimization scheme is employed for the determination of a linear projection of the training data (represented in the ELM space) to a low-dimensional feature space determined by the network target vectors. In this sense, ELM is closely related to kernel methods, like Kernel Spectral Regression (KSR) [1] and kernel Support Vector Machine (kSVM) [3]. Such methods by performing distance-based discrimination, are not able to highlight the most discriminative facial regions.

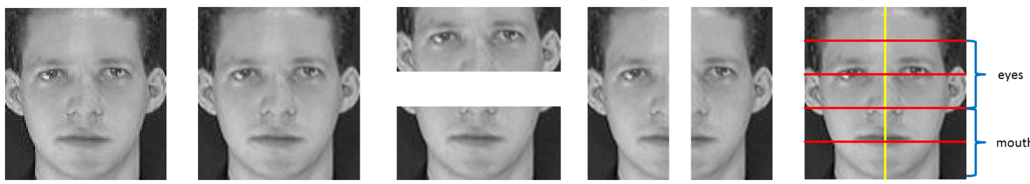
In this paper, we investigate the effectiveness of the ELM network in the facial image classification problem. Despite the fact that recent works in face recognition focus their attention on the unconstrained face recognition problem, usually referred to as 'Face Recognition in the Wild' [18, 19, 17, 2, 28], we focus our attention on a more restrictive application scenario requiring the classification of facial images depicting a nearly-frontal facial pose [14]. Such facial images can be obtained by face detection and tracking [21, 9, 29]. This problem needs to be addressed in many applications, e.g. applications facilitating assisted living of the elderly [27, 8], or relating to portable electronic devises usage. In such cases, the challenges that a facial image classification method needs to be able to face include illumination changes, different facial expressions and facial details (e.g., open/closed eyes), occluded facial regions (e.g., sun glasses and scarfs) and small face rotation/tilting (up to 20 degrees). We propose the exploitation of knowledge relating to the human face based on anthropometric ratios. We propose the use of multi-view Neural Networks, i.e., of multiple jointly-trained SLFN networks, each operating on a different facial region (view) in order to overcome the previously described weakness of the ELM algorithm. By optimally weighting the contribution of each view on the final classification result, the proposed classification scheme automatically combines discriminative information related to different facial regions, in order to enhance facial image classification performance. Experimental results on three publicly available databases denote that the combined approach is able to outperform facial image classification performed on pre-defined facial regions, as well as on the entire facial image. In addition, the adoption of the proposed facial view combination scheme outperforms other combination schemes that are usually employed in the literature.

The remainder of the paper is structured as follows. In Section 2 we describe the proposed facial image classification method. Section 3 presents experiments conducted in order to evaluate its performance. Finally, conclusions are drawn in Section 4.

2 Proposed Method

As has been previously described, we are interested in facial image classification in the cases where the depicted person has a nearly frontal facial pose. In Figure 1a we illustrate a facial image of the ORL database, having a frontal facial pose. Anthropometric ratios relating to the human face structure denote that it can be roughly partitioned in five equal horizontal zones. As can be seen in Figure 1b, the facial regions corresponding to the human eyes fall in the second and third zone, while the facial regions belonging to the mouth fall in the fourth and fifth zone. In addition, human faces are (in general) symmetric with respect to the vertical axis. As has been shown in [24], by exploiting the symmetry of the human face, enhanced facial image classification performance can be achieved. Thus, we can also roughly partition the facial images in two regions, as shown in Figure 1b.

In order to exploit the above-mentioned properties of human faces in an ELM-based classification framework, we propose the use of five facial image representations: the entire facial image, an image depicting the person's eyes, an image depicting the person's mouth and two images depicting the two (left and right half) partitions of the person's face with respect to the vertical axis, as illustrated in Figure 1c. Such representations can handle partial facial image occlusion that can create problems in face recognition [25] and facial expression recognition [20]. Furthermore, these facial facial image representations can be thought as different views of the facial image.



(a) (b) (c)

Figure 1: *a) Example of a human face having frontal facial pose, b) the same human face partitioned based on anthropometric ratios and c) the five facial image views used for face representation.*

After determining the facial views, we perform facial image classification based on multi-view Neural Networks. We employ the recently proposed Multi-view Regularized Extreme Learning Machine (MRELM) algorithm [12] to this end. A brief description of the MRELM algorithm is given in the following.

2.1 Multi-view Regularized Extreme Learning Machine

Let us denote by $\mathcal{U}$ a facial image database, containing N images, depicts N_P persons. We partition each facial image in order to create the above-described facial views. After facial view creation, each view image is represented as a matrix containing the (grayscale) intensity values and the obtained matrices are vectorized in order to produce vectors $\mathbf{x}_i^v \in \mathbb{R}^{D_v}$, $i = 1, \ldots, N$, $v = 1, \ldots, V$, where v runs along the different facial views and D_v is the dimensionality of the v-th facial view.

We would like to employ $\mathbf{x}_i^v$ and the corresponding labels c_i, in order to train V SLFN networks, each operating on one view. To this end we map the vectors of each view v to one ELM space $\mathbb{R}^{H_v}$, by using randomly chosen input weights $\mathbf{W}_{in}^v \in \mathbb{R}^{D_v \times H_v}$ and input layer bias values $\mathbf{b}^v \in \mathbb{R}^{H_v}$. H_v is the dimensionality of the ELM space related to view v. By doing this, each vector $\mathbf{x}_i^v$ is mapped to a vector $\boldsymbol{\phi}_i^v \in \mathbb{R}^{H_v}$. The network target vectors $\mathbf{t}_i = [t_{i1}, ..., t_{iN_P}]^T$, each corresponding to one facial image in the database, are set to $t_{ij} = 1$ for facial images depicting person j, i.e., when $c_i = j$, and to $t_{ij} = -1$ otherwise.

In order to determine both the network output weights $\mathbf{W}_{out}^v \in \mathbb{R}^{H_v \times N_P}$ and appropriate view combination weights $\gamma \in \mathbb{R}^V$, the following optimization problem is solved:

$$\textbf{Minimize:} \quad \mathcal{J} = \frac{1}{2}\sum_{v=1}^{V} \|\mathbf{W}_{out}^v\|_F^2 + \frac{c}{2}\sum_{i=1}^{N} \|\boldsymbol{\xi}_i\|_2^2 \tag{1}$$

$$\textbf{Subject to:} \quad \left(\sum_{v=1}^{V} \gamma_v \mathbf{W}_{out}^{v\,T} \boldsymbol{\phi}_i^v\right) - \mathbf{t}_i = \boldsymbol{\xi}_i, \;\; i = 1, ..., N, \tag{2}$$

$$\|\boldsymbol{\gamma}\|_2^2 = 1, \tag{3}$$

where $\boldsymbol{\xi}_i \in \mathbb{R}^{N_P}$ is the error vector related to the i-th facial image and c is a regularization parameter expressing the importance of the training error in the optimization process.

By setting the representations of $\mathbf{x}_i^v$ in the corresponding ELM space in a matrix $\mathbf{\Phi}^v = [\boldsymbol{\phi}_1^v, \ldots, \boldsymbol{\phi}_N^v]$, the network responses of the entire training set are given by:

$$\mathbf{O} = \sum_{v=1}^{V} \gamma_v \mathbf{W}_{out}^{v\,T} \mathbf{\Phi}^v. \tag{4}$$

By substituting (2) in (1) and taking the equivalent dual problem, $\mathcal{J}$ can be written as:

$$\mathcal{J}_D(\boldsymbol{\gamma}) = \frac{c}{2}\boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\gamma} - c\mathbf{r}^T \boldsymbol{\gamma} + \frac{\lambda}{2}\boldsymbol{\gamma}^T\boldsymbol{\gamma} + const, \tag{5}$$

where $\mathbf{P} \in \mathbb{R}^{V \times V}$ is a matrix having its elements equal to $[\mathbf{P}]_{kl} = tr\left(\mathbf{W}_{out}^{k\,T} \mathbf{\Phi}^k \mathbf{\Phi}^{l\,T} \mathbf{W}_{out}^l\right)$ and $\mathbf{r} \in \mathbb{R}^V$ is a vector having its elements equal to $\mathbf{r}_v = tr\left(\mathbf{T}^T \mathbf{W}_{out}^{v\,T} \mathbf{\Phi}^v\right)$. By solving for $\frac{\vartheta \mathcal{J}_D(\boldsymbol{\gamma})}{\vartheta \boldsymbol{\gamma}} = 0$, $\boldsymbol{\gamma}$ is given by:

$$\boldsymbol{\gamma} = \left(\mathbf{P} + \frac{\lambda}{c}\mathbf{I}\right)^{-1} \mathbf{r}. \tag{6}$$

By substituting (2) in (1) and taking the equivalent dual problem, $\mathcal{J}$ can also be written as:

$$\begin{aligned} \mathcal{J}_D(\mathbf{W}_{out}^v) &= \frac{1}{2}\sum_{v=1}^{V} tr\left(\mathbf{W}_{out}^{v\,T}\mathbf{W}_{out}^v\right) + \frac{c}{2} tr\left(\sum_{v=1}^{V}\sum_{l=1}^{V} \gamma_v \gamma_l \mathbf{W}_{out}^{v\,T} \mathbf{\Phi}^v \mathbf{\Phi}^{l\,T} \mathbf{W}_{out}^l\right) \\ &- c\sum_{v=1}^{V} tr\left(\gamma_v \mathbf{W}_{out}^{v\,T} \mathbf{\Phi}^v \mathbf{T}^T\right) + const. \end{aligned} \tag{7}$$

By solving for $\frac{\vartheta \mathcal{J}_D(\mathbf{W}_{out}^v)}{\vartheta \mathbf{W}_{out}^v} = 0$, $\mathbf{W}_{out}^v$ is given by:

$$\mathbf{W}_{out}^v = \left(\frac{2}{c\gamma_k}\mathbf{I} + \gamma_k \mathbf{\Phi}^v \mathbf{\Phi}^{v\,T}\right)^{-1} \mathbf{\Phi}^v (2\mathbf{T} - \mathbf{O})^T, \tag{8}$$

In order to jointly optimize $\mathcal{J}$ with respect to both $\mathbf{W}_{out}^v$ and $\boldsymbol{\gamma}$, an iterative optimization scheme formed by two optimization steps is followed, where $\mathbf{W}_{out,1}^v$ are initialized by training each SLFN network independently. The iterative optimization process is terminated when $(\mathcal{J}_D(t) - \mathcal{J}_D(t+1))/\mathcal{J}_D(t) < \epsilon$, where ϵ is a small positive value.

After the determination of the set $\{\gamma_v, \mathbf{W}_{out}^v\}_{v=1}^V$, the network response for a given set of facial view vectors $\mathbf{x}_t^v \in \mathbb{R}^D$ is given by:

$$\mathbf{o}_l = \sum_{v=1}^{V} \gamma_v \mathbf{W}_{out}^{v\,T} \boldsymbol{\phi}_t^v, \tag{9}$$

where $\boldsymbol{\phi}_t^v$ is the facial view vector representation in the ELM space corresponding to view v.

3 Experiments

In this Section, we describe experiments conducted in order to evaluate the performance of the proposed facial image classification scheme. We have employed three publicly available databases, namely the ORL, AR and Extended YALE-B databases. A brief description of the databases is given in the following Subsection. We have used the facial images provided by the databases and resized them to fixed size images of 40×30 pixels for computation speed consideration. Since there is not a commonly adopted training-test partitioning of the databases, we have randomly partitioned the databases in five sets, by keeping the 10%, 20%, 30%, 40% and 50% of the facial images depicting each person for training and the remaining facial images for testing.

In all the experiments we have employed the RBF activation function for the networks' hidden layer outputs calculation $\phi_{ik}^{v} = exp\left(\frac{\|\mathbf{x}_i^v - \mathbf{w}_k^v\|_2^2}{2\sigma_v^2}\right)$, where $\mathbf{w}_k^v$ is the weight of the k-th hidden layer neuron for view v. The value of σ_v was set equal to the mean Euclidean distance between the training vectors $\mathbf{x}_i^v$ and the network hidden layer weights $\mathbf{w}_k^v$. The dimensionality of the ELM space has been set to 1000 in all the cases, which is a value that has been shown to provide satisfactory performance in many classification problems.

3.1 Databases

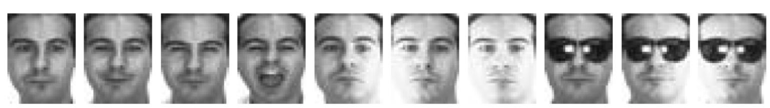
The ORL database contains 10 images of 40 persons, leading to a total number of 400 images ([26]). The images were captured at different times and with different conditions, in terms of lighting, facial expressions (smiling/not smiling) and facial details (open/closed eyes, with/without glasses). Facial images were taken in frontal position with a tolerance for face rotation and tilting up to 20 degrees. A set of ten images depicting a person of the database is illustrated in Figure 2a.

The AR database contains over 4000 images depicting 70 male and 56 female faces ([23]). In our experiments we have used the preprocessed (cropped) facial images provided by the database, depicting 100 persons (50 males and 50 females) having a frontal facial pose, performing several expressions (anger, smiling and screaming), in different illumination conditions (left and/or right light) and with some occluded facial regions (sun glasses and scarf). Each person was recorded in two sessions, separated by two weeks. A set of ten images depicting a person of the database is illustrated in Figure 2b.

The Extended YALE-B database contains images of 38 persons in 9 poses, under 64 illumination conditions ([22]) captured under varying lighting conditions. In our experiments we have used the frontal cropped images provided by the database. A set of ten images depicting a person of the database is illustrated in Figure 2c.

(a)

(b)



(c)

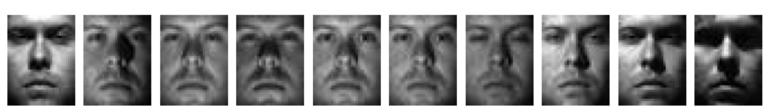


Figure 2: *Facial images depicting persons of the a) ORL, b) Extended YALE-B and c) AR datasets.*

3.2 Experimental Results

Tables 1, 2 and 3, illustrate the facial image classification rates obtained by applying the proposed approach in the ORL, AR and YALE-B databases, respectively. In these Tables we also provide the classification rates obtained by using each facial view independently and two commonly used view

Table 1: Classification rates on the ORL database.

	Face	Eyes	Mouth	FaceL	FaceR	Conc	Mean	**MV**
10%	73.06 %	55.28 %	59.17 %	68.89 %	62.78 %	71.94 %	71.39 %	**75.28** %
20%	81.56 %	61.56 %	63.75 %	74.69 %	74.38 %	87.81 %	84.06 %	**88.75** %
30%	85.71 %	65 %	69.29 %	80.71 %	78.93 %	88.57 %	87.14 %	**90.36** %
40%	88.33 %	69.17 %	73.33 %	80 %	85.42 %	91.67 %	90.83 %	**94.17** %
50%	87 %	71 %	76 %	81 %	83 %	93.5 %	91 %	**96** %

Table 2: Classification rates on the AR database.

	Face	Eyes	Mouth	FaceL	FaceR	Conc	Mean	**MV**
10%	39.52 %	39.91 %	28.26 %	41.17 %	40.7 %	49.52 %	47.65 %	**50.61** %
20%	40.71 %	47.95 %	36.33 %	38.95 %	45.48 %	50.86 %	52.38 %	**52.86** %
30%	56.06 %	**70.69** %	40.81 %	55.5 %	53.94 %	62.94 %	65.31 %	65.94 %
40%	58.94 %	**75.33** %	45.17 %	56.28 %	56.06 %	65.83 %	69.33 %	70.56 %
50%	77.23 %	81.08 %	49.96 %	76 %	74.54 %	87.31 %	88.85 %	**89** %

combination schemes, i.e. view combination by using the mean network output (noned by 'Mean') [10] and classification based on a combined data representation obtained by concatenating the facial view representations in the ELM space (noted by 'Conc'). It should be noted here that a facial view combination by concatenating the facial view representations in the input space is not expected to enhance facial image classification performance, when compared to the case of using only the entire facial image (i.e., the first facial view), since the same information is exploited in both cases.

As can be seen in these Tables, facial image classification by employing the entire facial image usually outperforms facial image classification based on the remaining facial regions. However, this is not always the case. In the AR database the facial region depicting the human eyes seems to be more discriminative, as the ELM network trained on this view outperforms the one trained on the entire facial image in all the cases. This fact may be explained by the fact that AR database contains facial images having occlusions (the persons wear sun glasses and scarfs).

Generally, view combination approaches increase the classification performance. Facial image classification based on a combined representation obtained by concatenating the view vectors generally

Table 3: Classification rates on the YALE-B database.

	Face	Eyes	Mouth	FaceL	FaceR	Conc	Mean	**MV**
10%	64.61 %	51.77 %	63.66 %	54.4 %	61.66 %	70.74 %	73.64 %	**74.64** %
20%	71.21 %	58.05 %	66.25 %	60.42 %	76.78 %	82.77 %	87.67 %	**89.78** %
30%	84.15 %	54.91 %	66.78 %	58.42 %	72.81 %	84.04 %	82.81 %	**88.01** %
40%	92.24 %	84.56 %	92.17 %	92.11 %	90.37 %	94.94 %	94.94 %	**95.43** %
50%	95.23 %	87.91 %	93.26 %	97.2 %	92.19 %	97.62 %	97.7 %	**97.86** %

outperforms single-view classification. Classification based on the mean network output outperforms the view concatenation approach in most cases. This seems reasonable, since each network trained on a different view tries to discriminate different facial representations independently. If we consider the network output vectors as a facial representation of increased discrimination power, the combination of discriminant facial image representations enhances classification performance.

Finally, the combination of the network outputs based on optimized combination weights further increases classification performance. It can be seen that the view combination scheme exploiting optimized weights outperforms the remaining view combination schemes in most cases. However, there are cases where this increase in the performance is marginal. This is observed in the AR and Yale-B databases when a relatively large training set is employed. In the cases where the adopted training set is small (e.g., in the case where the 10% of the data are used for training), the view combination approach determining optimized weights seems to be more effective.

Overall, it can be seen that the proposed facial image classification scheme outperforms the competing ones in most cases and achieves satisfactory performance in all the cases.

4 Conclusions

In this paper, we have investigated the effectiveness of the ELM network in (near-frontal) facial image classification. In order to enhance performance we proposed the use of multi-view ELM networks, which are trained in different facial representations (views) automatically obtained by exploiting anthropometric ratios. By jointly learning the network parameters and optimized network outputs combination weights, each facial region appropriately contributes to the final classification result. Experimental results on three publicly available databases show that the proposed approach outperforms facial image classification based on single facial views and on other view combination schemes.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART).

References

[1] D. Cai, X. He, and J. Han. Speed up Kernel Discriminant Analysis. *International Journal on Very Large Data Bases*, 20(1):21–33, 2011.

[2] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. *Computer Vision and Pattern Recognition*, 2013.

[3] R. E. Fan, P. H. Chen, and C. J. Lin. Working set selection using the second order information for training svm. *Journal of Machine Learning Researchy*, 6:1889–1918, 2005.

[4] T. Helmy and Z. Rasheed. Multi-category bioinformatics dataset classification using Extreme Learning Machine. *IEEE Evolutionary Computation*, 2009.

[5] G. B. Huang, L. Chen, and C. K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, 17(4):879–892, 2006.

[6] G. B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme Learning Machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 42(2):513–529, 2012.

[7] G. B. Huang, Q. Y. Zhu, and C. K. Siew. Extreme Learning Machine: a new learning scheme of feedforward neural networks. *IEEE International Joint Conference on Neural Networks*, 2004.

[8] A. Iosifidis, E. Marami, A. Tefas, I. Pitas, and K. Lyroudia. The MOBISERV-AIIA eating and drinking multi-view database for vision-based assisted living. *Journal of Information Hiding and Multimedia Signal Processing*, 6(2):254–273, 2015.

[9] A. Iosifidis, S. G. Mouroutsos, and A. Gasteratos. A hybrid static/active video surveillance system. *International Journal of Optomechatronics*, 5(1):80–95, 2011.

[10] A. Iosifidis, A. Tefas, and I. Pitas. Minimum Class Variance Extreme Learning Machine for human action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(11):1968–1979, 2013.

[11] A. Iosifidis, A. Tefas, and I. Pitas. Minimum Variance Extreme Learning Machine for human action recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014.

[12] A. Iosifidis, A. Tefas, and I. Pitas. Multi-view regularized Extreme Learning Machine for human action recognition. *Hellenic Conference on Artificial Intelligence*, 2014.

[13] A. Iosifidis, A. Tefas, and I. Pitas. Regularized Extreme Learning Machine for multi-view semi-supervised action recognition. *Neurocomputing*, 145:250–262, 2014.

[14] A. Iosifidis, A. Tefas, and I. Pitas. Class-specific Reference Discriminant Analysis with application in human behaviour analysis. *IEEE Transactions on Human-Machine Systems, D.O.I. 10.1109/THMS.2014.2379274*, 2015.

[15] A. Iosifidis, A. Tefas, and I. Pitas. Graph Embedded Extreme Learning Machine. *IEEE Transactions on Cybernetics, D.O.I. 10.1109/TCYB.2015.2401973*, 2015.

[16] A. Iosifidis, A. Tefas, and I. Pitas. On the kernel Extreme Learning Machine classifier. *Pattern Recognition Letters*, 54:11–17, 2015.

[17] R. Jafri and H. R. Arabnia. PCA-based methods for face recognition. *International Conference on Security & Management*, 2007.

[18] R. Jafri and H.R. Arabnia. A survey of component-based face recognition approaches. *International Conference on Artificial Intelligence*, 2007.

[19] R. Jafri and H.R. Arabnia. A survey of face recognition techniques. *Journal of Information Processing Systems*, 5(6):41–68, 2009.

[20] I. Kotsia, I. Buciu, and I. Pitas. An analysis of facial expression recognition under partial facial image occlusion. *Image and Vision Computing*, 26(7):1052–1067, 2008.

[21] K. Kucharski, W. Skarbek, G. Stamou, N. Nikolaidis, and I.Pitas. Morphological and adaboost face detectors comparison. *Workshop on Immersive Communication and Broadcast Systems*, 2005.

[22] K. C. Lee, J. Ho, and D. Kriegman. Acquiriing linear subspaces for face recognition under varialbe lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[23] A. Martinez and A. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.

[24] K. Papachristou, A. Tefas, and I. Pitas. Subspace learning with enriched databases using symmetry. *Euro-China Conference on Intelligent Data Analysis and Applications*, 2014.

[25] A. Rama, F. Tarres, L. Goldmann, and T. Sikora. More robust face recognition by considering occlusion information. *IEEE International Conference on Automatic Face & Gesture Recognition*, 2008.

[26] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. *IEEE Workshop on Applications of Computer Vision*, 1994.

[27] A. Tefas and I. Pitas. Human centered interfaces for assisted living. *International Conference on Man-Machine Interactions*, 2011.

[28] D. Yi, Z. Lei, and S. Z. Li. Towards pose robust face recognition. *Computer Vision and Pattern Recognition*, 2013.

[29] O. Zoidi, A. Tefas, and I. Pitas. Visual object tracking based on local steering kernels and color histograms. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(5):870–882, 2013.

[30] W. Zong and G. B. Huang. Face recognition based on Extreme Learning Machine. *Neurocomputing*, 74(16):2541–2551, 2011.