Journal of Naval Science and Engineering
2015, Vol. 11, No.3, pp. 1-13

KERNEL ELM BASED AGE ESTIMATION

Furkan GÜRPINAR[1], Heysem KAYA[2], Sadaf AFSHAR[3], Hamdi DİBEKLİOĞLU[4], Albert Ali SALAH[5]

[1,3]Program of Computational Science and Engineering, Boğaziçi University, Istanbul, Turkey
[2]Department of Computer Engineering, Namık Kemal University, Tekirdağ, Turkey
[4]Pattern Recognition & Bioinformatics Group, Delft University of Technology, Delft, Netherlands
[5]Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey

[1]furkan.gurpinar@boun.edu.tr, [2]heysem@boun.edu.tr, [3]sa.afshar.sa@gmail.com, [4]h.dibeklioglu@tudelft.nl, [5]salah@boun.edu.tr

Abstract

This paper presents a framework for estimating the apparent age of a subject from their face image. To learn the age estimation model, we used a set of visual descriptors and their feature-level fusion to obtain a single feature vector for apparent age. For model learning, we used regression with Extreme Learning Machines (ELM) with Gaussian (RBF) kernels. We tested the proposed system with k-fold cross-validation on the dataset provided by ChaLearn Looking at People 2015 - Apparent Age Estimation challenge. By combining visual descriptors from multiple grids, we obtained a Mean Absolute Error (MAE) of 5.20.

Keywords*: Age estimation, ELM, biometrics*

1. INTRODUCTION

Automated age estimation from facial imagery is one of the most difficult challenges in face analysis. It can be very useful in a number of

application areas such as video surveillance, authorization systems, demographic data mining and business intelligence. The difficulty of this task is due to many reasons, which include the limitations in data collection such as the lack of enough samples to model the aging patterns of subjects, as well as uncontrolled data in terms of lighting, pose and other environmental variables. Aging process is also known to be very subject-dependent, i.e. subjects might differ in terms of aging patterns, which results in high variations within the samples from the same age. One of the earliest works regarding automated age estimation from faces is done in early 2000s by Lanitis et al. [16, 17]. After large databases such as MORPH [25], FRGC [23] and FG-NET [1] became available, the interest on this topic has grown significantly, and many different feature extraction and learning schemes have been tried for the task of age estimation from faces. We provide a brief literature review on these works in the following section.

1.1. Related Work

Various feature extraction methods were applied in order to model the aging patterns of subjects from their facial imagery. For example, Active Shape Models (ASM) and Active Appearance Models (AAM) have been used as visual cues for age estimation [3, 9, 16, 19]. Bio-Inspired Features (BIF) [26] is another feature extraction technique for age estimation, which is consistently used in recent years [14]. BIF features summarize images with a multi-layer feed-forward model where the first layer convolves the image with a set of Gabor filters from multiple orientations and scales. Then a pooling step -usually with STD or MAX operators- downsizes the resulting vector. In [13], the authors use a simplified version of this model by setting the number of bands and orientations manually. Histogram-based local appearance features have also been used for age estimation, such as the Local Binary Patterns (LBP) descriptor in combination with PCA projections of both the BIF descriptor and the original image pixels in [28], and Histogram of Oriented Gradients (HOG) descriptor in [7]. Regarding the learning schemes, many algorithms have been employed for the age estimation task. For example, Support Vector Machine Regression (SVR) is

a commonly used algorithm for this task [3, 13, 14,28]. Support Vector Machines (SVM) are also used with modifications such as a binary tree where each node has an SVM in [14], and alternative ranking formulations of support vectors such as OHRank [3] and Multi-feature ordinal ranking [28].

Other algorithms regarding this task include Random Forests [20], and Neural Network architectures [8, 16], as well as projection based learners such as Partial Least Squares regression and Canonical Correlation Analysis, which are oftenly used in combination with kernel and regularization techniques [10, 11, 12]. This study is conducted on the data set provided by ChaLearn Looking at People 2015 - Apparent Age Estimation challenge. A variety of methods were implemented by different participants, and it's found that there is still room for improvement for any part of the pipeline, namely the face detection/alignment, feature extraction and model learning. For learning, almost all top ranking participants used external datasets to increase the number of training samples. Convolutional Neural Networks (CNN) are also commonly used by the top ranking participants, including the winner of the challenge.

2. METHODOLOGY

In this section we briefly describe the steps of the proposed system, which are face registration, feature extraction and age estimation.

2.1. Face Registration

Before extracting visual features, we align the faces by detecting the locations of landmark points by using Xiong and De la Torre's Supervised Descent Method (SDM) [29]. After the landmarks are located, the face is cropped from a bounding box around the outer landmarks, and it's resized to 64 x 64 pixel images, and the registered faces' histograms are equalized in order to deal with the variations in skin color and illumination. Due to the high number of rotated images in the data set, we also rotate the original images in order to increase the detection rate. A summary of the face

detection rates is supplied in Table 1.

2.2. Visual Descriptors

We extract visual descriptors such as SIFT, HOG, LBP and Gist from local patches, which are obtained by dividing each image into regular, non-overlapping blocks. We extracted the features from 8, 16 and 32 pixel patches.

1) *Local Binary Patterns:* After aligning the faces and converting to grayscale, we extracted LBP feature which encodes a 3 x 3 pixel neighborhood with an 8-bit number depending on the binary relations of each pixel with the central pixel's intensity value. Therefore the LBP descriptor represents a local region with a 256-bin histogram. Using only 58 uniform patterns and 1 additional bin for the non-uniform ones, the histogram becomes 59-dimensional per patch, which represents the frequencies of common patterns [21].

2) *Histogram of Oriented Gradients:* Aligned faces are also processed with the HOG feature extraction method, which represents each local region by a histogram of edge orientations [4]. We use the HOG variant in [6] which results in a 31-dimensional feature per patch.

3) *Scale-invariant Feature Transform:* SIFT features summarize an interest point of a grayscale image using the statistics of gradient directions of the intensity levels of an image [18]. Instead of detecting the interest points, we chose to extract SIFT features from regular grid points, which results in a 128-dimensional feature vector per grid point.

Apart from these descriptors, we also tested the performances of various descriptors such as the GIST descriptor, which produces a simplified visual summary of a given image [27]. We also tried the LGBP descriptor, which is LBP extracted from the convolutions of the original image with a set of Gabor filters, as explained in [2], and the LPQ descriptor

which is known to be blur insensitive and informative in many face-related tasks [22]. Finally, we employ a set of geometric features that summarize the distances and angles between sets of stationary landmark points, such as the nose and eyes. These features are extracted using the landmarks detected by SDM. The individual performances of classifiers that use these features are displayed in Table 2.

2.3. Model Learning

In order to learn a regression model, we used kernel Extreme Learning Machine (ELM) [15] due to its learning speed and accuracy. In the following paragraphs, we briefly explain the working principle of ELMs. ELM proposes a simple and robust learning algorithm in single-hidden layer feedforward networks (SLFNs), which is visualized in Figure 1. The input layer is initialized with a random bias and set of weights, to obtain the output of the second (hidden) layer. After the hidden layer's output is calculated, the set of weights and the bias term that link the hidden layer to the output layer (i.e. the label vector) is calculated analytically by a simple generalized inverse operation of the hidden layer output matrix. This calculation is explained in more detail in the following paragraph. ELM aims to map the hidden node output matrix $\mathbf{H} \in \mathbf{R}^{N\,x\,h}$ to the label vector $\mathbf{T} \in \mathbf{R}^{N\,x\,1}$, where N and h denote the number of samples and the hidden neurons, respectively. The set of output weight s$\beta \in \mathbf{R}^{h\,x\,1}$ is learned by the least squares solution of the set of linear equations Hβ = T, as:

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}\,, \tag{1}$$

where H† represents the Moore-Penrose generalized inverse [24] that minimizes the L2 norms of ||Hβ -T|| and || β || simultaneously. To increase the robustness and the generalization capability of ELM, a regularization coefficient C is added to the optimization procedure. Therefore, given a kernel K, the set of weights is learned as follows :

$$\beta = \left(\frac{\mathbf{I}}{\mathbf{C}} + \mathbf{K}\right)^{-1}\mathbf{T}. \tag{2}$$

In order to calculate kernel K from the original features, we choose to use the Gaussian (RBF) kernel because its performance is found to be superior compared to the alternatives such as linear and polynomial kernels.

2.4. Fusion

To combine high-dimensional feature vectors, we chose to perform Principal Component Analysis (PCA) prior to feature-level fusion. In order to get the most varying features from each patch, we learn the projections that account for the 90% of the variance for each patch independently, then combine the projected vectors to obtain the reduced set of one feature. We apply this method for all features and concatenate the resulting feature vectors. The pipeline that summarizes the proposed age estimation system is given in Figure 2.

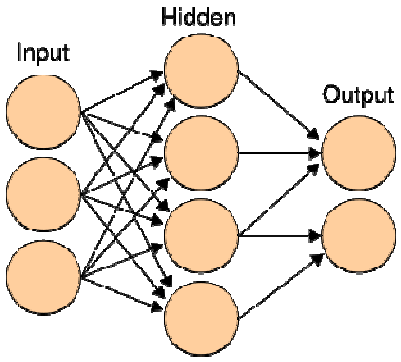

Figure 1. Single-hidden layer feedforward neural network architecture of ELM.

3. EXPERIMENTAL SETUP

In this section, we briefly explain the dataset used for training and testing the system, and we display our experimental results.

3.1. Corpus

For testing the performance of our system, we used the dataset provided by ChaLearn Looking at People 2015 Apparent Age Estimation challenge. The challenge corpus contains images of people labeled with the mean and standard deviations of human annotators' labels for their apparent ages [5]. The dataset is partitioned into training, validation and test sets. The distribution of samples over the partitions is given in Table 1.

Table 1. Number of images per partition. First line: Given data, Second line: Number of aligned faces by our system

#	Train	Val	Test
Given	2476	1136	1079
SDM	2311	1050	1010

3.2. Performance Measures

1) Mean Absolute Error (MAE) is a standard way of measuring the performance of a continuous predictor, where the accuracy is measured in terms of the magnitude of the deviation from the true value. MAE scores of each test sample are averaged to obtain the final performance of the predictor. So MAE measure can be summarized as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i|, \tag{3}$$

where $\mathbf{x_i}$ is the prediction, and $\hat{\mathbf{x}}_i$ is the true label, i.e. the average of annotations for sample i.

2) Challenge Score (CS): Since the data provided by ChaLearn Looking at People 2015 - Apparent Age Estimation Challenge is labeled by multiple human annotators for the subjects' apparent age, the performance of a predictor is more accurately measured when the standard deviation of the annotations is also taken into account. So the challenge score is calculated as follows:

$$\varepsilon = 1 - e^{\frac{(x-\mu)^2}{2\sigma^2}}, \quad (4)$$

wherex is the estimated age, μ and σ are the mean and standard deviation of the annotations for that sample, respectively. The overall score is the average error terms over all samples in the testing set, which is $\in [0,1]$.

3.3. Results

To obtain the optimal settings for the predictor, we optimized the hyperparameters of the model such as the RBF kernel parameter γ, regularization parameter C for ELM and number of PCA eigenvectors p, using 3-fold cross-validation. Using these parameters, we combine the projected features from each visual descriptor, and learn the regression model using this fused training set. Performances of various descriptors with the best grid parameters are given in Table II.

Using the best grid parameters for the 4 visual descriptors (Gist, HOG, LBP and SIFT), we concatenate each feature after reducing with PCA to keep 90% variance, to obtain the final feature vector, the performance of which is given in Table III. Since the test set labels are sequestered, we have MAE evaluations for only the training and validation sets. We also display some examples from the validation set in Figure 3. It can be seen from the bottom row that the failure cases usually correspond to blur and occlusions, which are common problems in many face-related applications.

Table 2. Performance of various descriptors extracted from aligned images

Descriptor	**MAE**	**CS**	**Dimensions**
LBP	5.63	0.49	3712
HOG	5.64	0.49	4464
SIFT	5.64	0.49	3200
GIST	5.42	0.47	2048

LPQ	6.62	0.55	9216
LGBP	7.59	0.6	16704
GEO	8.54	0.64	18

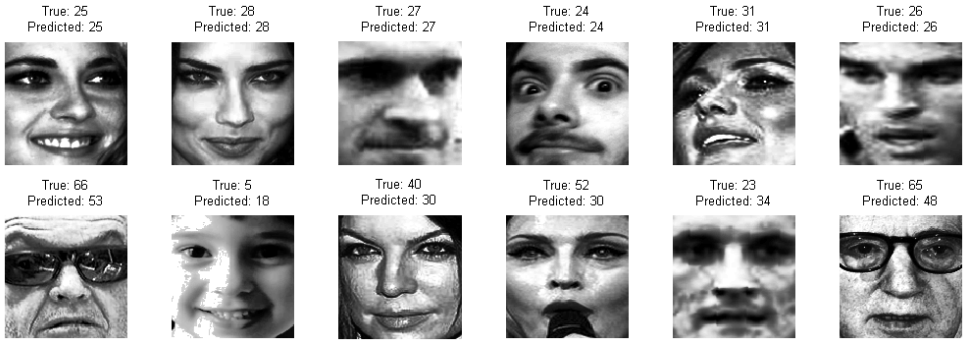

Figure 1. Samples from the validation set. Top row are correct predictions, bottom row are the failure cases. Blurring and occlusions often cause errors in prediction.

Table 3. Performance of the proposed fusion on data partitions

Dimensions	**MAE-Train**	**MAE-Val**	**CS-Train**	**CS-Val**	**CS-Test**
13302	5.20	5.44	0.46	0.48	0.52

4. CONCLUSIONS

In this study, we tested several appearance descriptors that are extracted from local regions of facial images. We selected the best performing ones and combined them in order to obtain a combined apparent age feature vector. We then used a least-squares based regressor to learn a model that maps this combined feature set to the apparent age of the people, and we saw that this combination yields significantly better results compared to the individual performances of these descriptors.

As future work, we want to analyze age estimation systems with the use of dynamic data, i.e. videos of people, which are not considered in this

study. This way, we will be able to keep the appearance information, and add the information related to the motion of facial muscles that can be informative in determining a person's age [30]. Moreover, interaction, interpersonal synchrony, and mimicry in two-party communications have a lot of implications. This kind of data, particularly from sessions in psychotherapy or consulting (with either symmetrical or asymmetrical relationships) could be very interesting to analyze.

REFERENCES

[1] The FG-NET aging database, November 2002. Available at http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm.

[2] T. R. Almaev and M. F. Valstar. Local Gabor binary patterns from three orthogonal planes for automatic facial expression recognition. In *Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 356–361. IEEE, 2013.

[3] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 585–592. IEEE, 2011.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893. IEEE, 2005.

[5] S. Escalera, J. Gonzalez, X. Bar´o, P. Pardo, J. Fabian, M. Oliu, H. Escalante, I. Huerta, and I. Guyon. ChaLearn 2015 apparent age and cultural event recognition: datasets and results. In *ICCV ChaLearn Looking at People Workshop*, 2015.

[6] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627– 1645, 2010.

[7] C. Fernandez, I. Huerta, and A. Prati. A comparative evaluation of regression learning algorithms for facial age estimation. In *Face and Facial Expression Recognition from Real World Videos*, pages 133–144. Springer, 2015.

[8] X. Geng, C. Yin, and Z.-H. Zhou. Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2401–2412, 2013.

[9] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2234–2240, 2007.

[10] G. Guo and G. Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 657–664. IEEE, 2011.

[11] G. Guo and G. Mu. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In 10th *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1–6. IEEE, 2013.

[12] G. Guo and G. Mu. A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, 32(10):761– 770, 2014.

[13] G. Guo, G. Mu, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 112–119. IEEE, 2009.

[14] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In *International Conference on Biometrics*, pages 1–8. IEEE, 2013.

[15] 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.

[16] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):621–628, 2004.

[17] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[19] K. Luu, K. Ricanek Jr, T. D. Bui, and C. Y. Suen. Age estimation using active appearance models and support vector machine regression. In *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–5. IEEE, 2009.

[20] A. Montillo and H. Ling. Age regression from faces using random forests. In 16th *IEEE International Conference on Image Processing*, pages 2465–2468. IEEE, 2009.

[21] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971– 987, 2002.

[22] V. Ojansivu and J. Heikkila. Blur insensitive texture classification using local phase quantization. In *Image and signal processing*, pages 236– 243. Springer, 2008.

[23] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE computer society conference on Computer vision and pattern recognition*, pages 947–954, 2005.

[24] C. R. Rao and S. K. Mitra. *Generalized inverse of matrices and its applications*, Wiley, New York, 1971.

[25] K. Ricanek Jr and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 341–345. IEEE, 2006.

[26] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.

[27] A. Torralba, K. P. Murphy, W. T. Freeman, M. Rubin, et al. Contextbased vision system for place and object recognition. In *9th IEEE International Conference on Computer Vision*, pages 273– 280, 2003.

[28] R. Weng, J. Lu, G. Yang, and Y.-P. Tan. Multi-feature ordinal ranking for facial age estimation. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1–6. IEEE, 2013.

[29] X. Xiong and F. De la Torre. Supervised Descent Method and Its Application to Face Alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 532–539, 2013.

[30] H. Dibeklioglu, F. Alnajar, A.A. Salah, and T. Gevers. Combining Facial Dynamics With Appearance for Age Estimation. *IEEE Transactions on Image Processing*, 24(6): 1928--1943, 2015.