LOCAL AGE GROUP MODELING IN UNCONSTRAINED FACE IMAGES FOR FACIAL AGE CLASSIFICATION

*Seung Ho Lee and Yong Man Ro**

Department of EE, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea

ABSTRACT

Age classification in the real world is a very challenging task due to the large variation of face appearances (e.g., a variety of human races, genders, facial expressions, poses etc.). In this paper, we propose a new age classification method using local modeling of age group to deal with large variation problem. The local modeling is built by clustering training faces within an age group. Nearest face clusters in the local modeling to a test face contribute in determining the age group of the test face. This enables us to reduce the effect of the variation unrelated to age. For comparing the test face with the face clusters, we combine two complementary similarities that consider the cluster centroid and the intra-cluster variation. Experimental result on a real-world dataset shows that our local modeling based approach is superior to global modeling based approach, achieving state-of-the-art performance.

Index Terms**—**Facial age classification, local age group model, face clustering

1. INTRODUCTION

Recently, automatic age classification using face information has gained an increasing interest due to its emerging applications such as forensic art, electronic customer relationship management (ECRM), security control and surveillance monitoring, and biometrics [1].

So far, many age classification methods have been proposed. In the beginning, most of the age classification methods are based on facial landmark localization [2-4][8-10]. Kwon and Lobo proposed to compute six ratios of distances between facial components (e.g., eye, nose, mouth etc.) and used them to separate babies from adults [2][3]. Lanitis et al. [4] adopted an active appearance model (AAM) [6] to build a shape and intensity model from a set of training face images. Using the built shape-intensity model, a new test image was parameterized and classified as an actual age by an age function learned with a genetic algorithm [4]. The AAM was also adopted in many other age classification methods [8-10]. However, facial landmark localization based approaches might not be proper for face images taken under uncontrolled conditions. The accuracy of localizing face landmarks could significantly drop when the face images are of low-resolution or occluded [7]. Indeed, the experiments of [2-4][8-10] were performed on relatively limited data sets with well-controlled conditions (e.g., high resolution face images without occlusion).

Since Gallagher and Chen [11] have recently constructed a large dataset of faces (named Images of Groups database) collected from Flickr [12], a few studies have been conducted to address the age classification under uncontrolled conditions using the Images of Groups database. In [13], two widely used appearance features, local binary patterns (LBPs) [14] and Gabor wavelets [15], were used for representing faces. Furthermore, Adaboost [16] was adopted to select the discriminative features (e.g., LBP histogram bins) and combined them into a strong classifier or used them as an input to a nonlinear support vector machine (SVM) [20] classifier. In [18], LBP variants that contained micro patterns and its strength information were used for face representations, followed by a nonlinear SVM classifier. Note that the methods in [13][18] globally modeled faces of age groups (i.e., representing each age group as a single model). Globally modeling age groups was problematic in an uncontrolled condition. This is because faces of same age group could be significantly different or faces in different age groups could be similar due to the large face appearance variations. The large variations are mainly due to a variety of facial attributes, such as identity, race, gender, facial expression, pose etc. [1]. Moreover, it is likely that Adaboost and nonlinear SVM are susceptible to overfitting training set [19][20].

In this paper, we propose a new age classification method suitable for dealing with the large face appearance variations. The main contribution of this paper is twofold.

1) For the proposed local modeling, we decompose the whole training faces of an age group into a set of face clusters. Using a few nearest face clusters of each age group, we can avoid the classification degradation due to some undesirable faces (or outliers).
2) We propose an effective way for measuring the closeness between a test face and a face cluster. We compute the distance between the test face and the face cluster centroid that represents the representative face appearance. To further reflect the variation of face appearance, we compute the distance between a test face and the face sample distribution formed with the training faces in the face cluster. These two complementary measures are useful for improving the separability between face clusters from different ages.

Comparative experiments have been conducted using the Images of Groups database that contains Flickr images taken under uncontrolled environments. Experimental results show

* Corresponding author (ymro@ee.kaist.ac.kr)

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

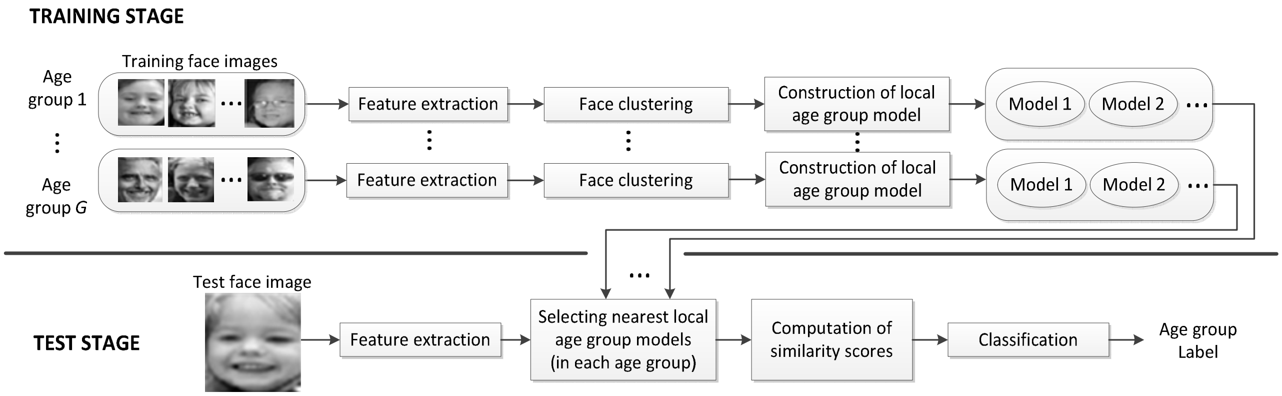

Fig. 1. Overview of the proposed age classification based on local age group modeling.

that the proposed age classification using local age group modeling outperforms other state-of-the-art approaches using global age group modeling.

The rest of this paper is organized as follows: Section 2 details the proposed age classification based on the local age group modeling. Section 3 presents the experimental results and conclusions are drawn in Section 4.

2. PROPOSED AGE CLASSFICATION USING LOCAL AGE GROUP MODELING

Fig. 1 shows the overview of the proposed age classification method. The training stage (i.e., construction of local age group models from training faces) consists of feature extraction, face clustering, and local age group model construction. The aim of feature extraction is to capture rich age-related information from face image. In the feature space, we apply a face clustering to the training samples so that faces with similar appearance can be grouped together. Next, each face cluster is converted into a local age group model that enables matching with test face image in test stage.

In test stage (i.e., classification of test face using local age group models), the feature vector of a test face is compared with every local age group model (generated in the offline training stage). From the comparisons, we select the few local age group models nearest to the test face, in each age group. Next, we compute the overall similarity scores (each of which corresponds to an age group) using the selected local age group models. Finally, the age group of the test face is determined by finding the highest similarity score among all age groups.

Detailed descriptions of the proposed local age group model construction and classification are given in the Section 2.1 and Section 2.2, respectively.

2.1. Construction of local age group models from training faces

In this section, we describe the construction of the proposed local models for a given age set of training face images. The feature vector is extracted from each age training face image. In this paper, we use local binary pattern (LBP) histogram [14] as feature vector since it can efficiently encode facial micro-patterns [14] related with age. Because LBP feature extraction leads to a high dimensional feature vector, a dimension reduction technique is applied to the LBP features. For the dimension reduction technique, we adopt Fisher's linear discriminate analysis (FLDA) [21] (that has been widely used in face analysis) to reduce computational time and maintain discrimination capability in classification. For a total of G age groups, FLDA generates a $G-1$ dimensional feature vector [21]. It should be noted that any other feature extraction or dimension reduction can be utilized with the proposed method.

For constructing local age group models by using the feature vectors of the training face images, let $\mathbf{F}^{(i)} = \left[\mathbf{f}_1^{(i)}, \mathbf{f}_2^{(i)}, ..., \mathbf{f}_{N_i}^{(i)}\right] \in \Re^{(G-1)\times N_i}$ denotes the training set of the i-th $(i = 1,...,G)$ age group, which consists of N_i feature vectors denoted by $\mathbf{f}_n^{(i)} \in \Re^{G-1}$. We aim to divide $\mathbf{F}^{(i)}$ into a set of face clusters where each has similar characteristics (in terms of gender, races, facial pose, hair style etc.). Because the number of face clusters may vary with different datasets, a hierarchical clustering method seems more appropriate than clustering methods with a pre-defined number of clusters (such as k-means clustering [23]). In this paper, hierarchical agglomerative clustering (HAC) [24] is adopted as the clustering method. In HAC, the number of face clusters is automatically determined by cutting the dendrogram (which is a tree diagram that represents all the ways to one single large cluster [17]) at a specific similarity level [17]. From the clustering result, the set of face clusters for the i-th age group can be defined as $\mathbf{S}^{(i)} = \left\{\mathbf{C}_k^{(i)}\right\}_{k=1}^{K_i}$ where $\mathbf{C}_k^{(i)}$ is the k-th face cluster in $\mathbf{S}^{(i)}$. Note that performing the clustering within individual age groups ensures that feature vectors from different age groups are not included in same face cluster.

We represent the k-th face cluster as $\mathbf{C}_k^{(i)} = \left[\mathbf{f}_{k,1}^{(i)}, \mathbf{f}_{k,2}^{(i)}, ..., \mathbf{f}_{k,N_{i,k}}^{(i)}\right] \in \Re^{(G-1)\times N_{i,k}}$, where $\mathbf{f}_{k,m}^{(i)}$ is the m-th feature vector contained in $\mathbf{C}_k^{(i)}$ and $N_i = \sum_{k=1}^{K_i} N_{i,k}$. To obtain the representative face appearance information, we compute the cluster centroid ($\bar{\mathbf{f}}_k^{(i)}$) of $\mathbf{C}_k^{(i)}$ as

$$\bar{\mathbf{f}}_k^{(i)} = \frac{1}{N_{i,k}} \sum_{m=1}^{N_{i,k}} \mathbf{f}_{k,m}^{(i)}. \quad (1)$$

In order to additionally encode sample variation information in $\mathbf{C}_k^{(i)}$, we make use of principal component analysis (PCA) [22] which seeks a linear subspace (or a set of basis vectors) that maximizes total variance of samples. The linear subspace formed with the feature vectors in the face cluster $\mathbf{C}_k^{(i)}$ is defined as $\mathbf{B}_k^{(i)} = [\varphi_1, \varphi_2, ..., \varphi_M] \in \Re^{(G-1)\times M}$ where φ_m is the basis vector of $\mathbf{B}_k^{(i)}$ that corresponds to the m-th largest eigenvalue [22], and $M \le G-1$.

Using $\bar{\mathbf{f}}_k^{(i)}$ and $\mathbf{B}_k^{(i)}$, the local age group model ($\mathbf{L}_k^{(i)}$) for the face cluster $\mathbf{C}_k^{(i)}$ is as follows:

$$\mathbf{L}_k^{(i)} = (\bar{\mathbf{f}}_k^{(i)}, \mathbf{B}_k^{(i)}). \quad (2)$$

Once every local age group model $\mathbf{L}_k^{(i)}$ ($i = 1,...,G$ and $k = 1,...,K_i$) is generated from the associated face cluster, age classification for test face image can be performed, which will be detailed in the next subsection.

2.2. Classification of test face by using local age group models

In this section, we describe the age classification of a given test face image by using the local age group model of (2).

Let us define the feature vector of the test face image by $\mathbf{q} \in \Re^{G-1}$. Note that $\mathbf{q}$ is obtained using similar way to the feature extraction of the training face images (each of which is denoted by $\mathbf{f}_n^{(i)}$). For the comparison of $\mathbf{q}$ and local age group model, we define the closeness between them. The closeness between $\mathbf{q}$ and local age group model $\mathbf{L}_k^{(i)}$ is computed as the distance $d_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)})$ between $\mathbf{q}$ and the cluster centroid $\bar{\mathbf{f}}_k^{(i)}$ of $\mathbf{L}_k^{(i)}$ as follows:

$$d_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)}) = \left\| \mathbf{q} - \bar{\mathbf{f}}_k^{(i)} \right\|_2, \quad (3)$$

where $\| \cdot \|_2$ is ℓ_2-norm operator. In order to reflect variation of face appearance for the computation of closeness to $\mathbf{L}_k^{(i)}$, we compute the distance $d_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)})$ between $\mathbf{q}$ and the linear subspace $\mathbf{B}_k^{(i)}$ of $\mathbf{L}_k^{(i)}$. This can be achieved by computing the PCA residual [26] between $\mathbf{q}$ and the reconstruction of $\mathbf{q}$ projected onto the linear subspace $\mathbf{B}_k^{(i)}$ as follows:

$$d_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)}) = \left\| \left(\mathbf{q} - \bar{\mathbf{f}}_k^{(i)}\right) - \mathbf{B}_k^{(i)} \mathbf{B}_k^{(i)\mathrm{T}} \left(\mathbf{q} - \bar{\mathbf{f}}_k^{(i)}\right) \right\|_2, \quad (4)$$

where the term $\mathbf{B}_k^{(i)} \mathbf{B}_k^{(i)\mathrm{T}} \left(\mathbf{q} - \bar{\mathbf{f}}_k^{(i)}\right)$ corresponds to the reconstruction of $\mathbf{q}$ (subtracted by the cluster centroid $\bar{\mathbf{f}}_k^{(i)}$). Fig. 2 shows a conceptual illustration for the computation of the closeness between a test face and the local age group models. In the example shown in Fig. 2, we can see that the

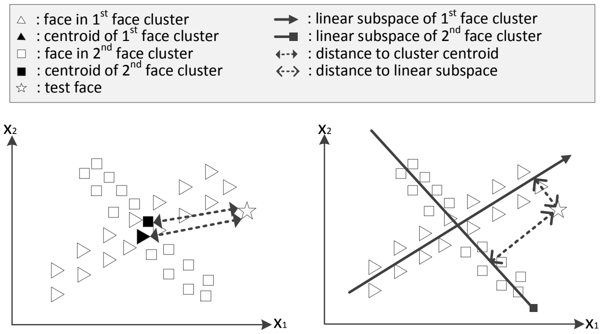

Fig. 2. Conceptual illustration on two-dimensional space to show the usefulness of computing closeness between test face and local age group model by considering variation of face appearance.

distances from the linear subspaces could provide information complementary for discriminating two local age group models when the distances from the representative face appearances (i.e., cluster centroids) of the two local models are similar. Before combining the two Euclidean distances $d_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)})$ and $d_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)})$, we normalize so that they have the same scale. To this end, we convert the distances into the similarity scores $s_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)})$ and $s_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)})$ (that range from 0 to 1) by using radial basis function (RBF) [25]:

$$s_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)}) = \exp(-d_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)}) / \beta). \quad (5)$$

$$s_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)}) = \exp(-d_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)}) / \beta). \quad (6)$$

In (5) and (6), β is a weight scale parameter. These two similarity scores are fused together using weighted sum:

$$s(\mathbf{q}, \mathbf{L}_k^{(i)}) = w \cdot s_{\text{cc}}(\mathbf{q}, \mathbf{L}_k^{(i)}) + (1-w) \cdot s_{\text{ss}}(\mathbf{q}, \mathbf{L}_k^{(i)}), \quad (7)$$

where w denotes the weight value that varies within [0, 1]. In (7), the former term represents the closeness to the representative face sample appearance and the latter reflects the closeness to the face sample distribution in the cluster $\mathbf{L}_k^{(i)}$.

For classification of age groups, we compute the overall similarity score $\bar{s}(\mathbf{q}, \mathbf{S}^{(i)})$ for each age group, as the average of the R largest similarity scores among all $s(\mathbf{q}, \mathbf{L}_k^{(i)})$ ($k = 1,...,K_i$) where $R \le K_i$. Note that, when R is too small (e.g., $R = 1$), the classification may overfit the training data. On the other hand, a large R (e.g., $R = K_i$) may also degrade the classification performance by incorporating the training face images that are very different from the test face image. Finally, the age group label ℓ of the feature vector $\mathbf{q}$ of the test face image is determined by finding the maximum similarity score among G age groups as follows:

$$\ell = \arg\max_{i=1}^{G} \left(\bar{s}(\mathbf{q}, \mathbf{S}^{(i)}) \right). \quad (8)$$

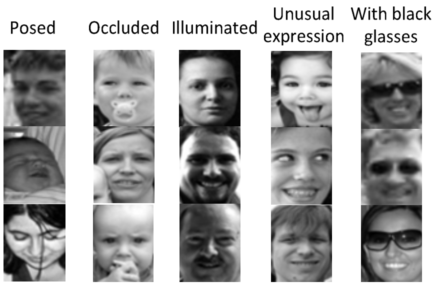

Fig. 3. Example face images from the Images of Groups database.

3. EXPERIMENT

To evaluate the proposed age classification method, the Images of Groups database [11] were used. This database contained 28,231 faces (from 5,080 Flickr images) each of which was labeled as one of the seven age groups: 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+ [11]. Faces often had low resolution (e.g., 25% of the faces have under 12.5 pixels between the eye centers), occlusion, dark glasses, unusual facial expressions etc. [11]. We cropped the face regions and normalized them to 60 (height) x 48 (width) pixel images using the eye center positions provided by the authors in [11] (see Fig. 3).

For a fair comparison, we followed the benchmarking protocol used in [11][13][18]. Specifically, we randomly selected 3,500 (i.e., 500 per age group) face images as the training set, and the remaining 1,050 (i.e., 150 per age group) face images were used as the test set. This dataset was termed as the 'Full set'. To study age classification on faces with reasonable resolution, we collected 12,080 face images with the eye distance more than 24 pixels [18][27]. Among these, following [18][27], we selected 2,080 face images (300 per age group except for the age group '8-12' in which only 280 face images were available [18]) as the training set, and 664 face images (100 per age group except for the age group '8-12' in which only 64 face images were available [18]) as the test set. This dataset was termed as the 'More than 24 pixels'.

For feature extraction, each face image was divided into 6 x 6 non-overlapping local regions. Using an LBP operator with (8,1) neighborhood [14], LBP histogram was extracted from each of the 80 (=10 x 8) local regions, and the local regional LBP histograms were concatenated to form an LBP histogram of the face image. In order to generate the projection matrix used in the FLDA, the whole face images in the associated training set were used for each age group. In HAC clustering, we set the cutting point to 65% of the dendrogram height, which produced 7~12 face clusters in each age group. In experiment, we set the weight scale parameter β in (5) and (6) to 20. The weight parameter (w) in (7) and the number of nearest face clusters used for determining age group (R) were set to 0.8 and 4, respectively. The number of basis vectors (M) contained in linear subspace $\mathbf{B}_k^{(i)}$ was set to 4.

To investigate the effect of decomposing each age group into several local age models, we evaluated the proposed method and its baseline in which face clustering was not applied. Specifically, in the baseline, K_i (i.e., the number of face clusters) was set to 1 (global age modeling) for $i = 1,...,G$ while the other conditions such as feature extraction and classification remained the same as the proposed method. We also made comparisons with state-of-the-art performances in [11][13][18][27] under the same protocol.

Table 1. Comparisons with the baseline and the state-of-the-arts.

Age classification method	Classification accuracy	
	Full set	**More than 24 pixels**
Appearance [11]	38.3 %	Not available
Appearance + Context [11]	42.9 %	Not available
Gabor + Adaboost [13]	43.7 %	48.2 %
LBP + Adaboost [13]	44.9 %	48.3 %
Boosted Gabor + SVM [13]	48.4 %	52.6 %
Boosted LBP + SVM [13]	50.3 %	55.9 %
LBP variants + SVM [18]	51.7 %	Not available
Soft encoding [27]	Not available	59.5 %
Baseline	45.5 %	51.1 %
Proposed method	**53.2** %	**59.8 %**

Table 2. Confusion matrix using the proposed method (For full set). Row and column mean actual and predicted age groups

	0-2	**3-7**	**8-12**	**13-19**	**20-36**	**37-65**	**66+**
0-2	**75 %**	16 %	6 %	2 %	1 %	0 %	0 %
3-7	8 %	**50** %	29 %	7 %	3 %	3 %	0 %
8-12	2 %	19 %	**46** %	20 %	9 %	3 %	1 %
13-19	0 %	7 %	16%	**35** %	29 %	9 %	4 %
20-36	1 %	2 %	7 %	17 %	**51** %	17 %	6 %
37-65	0 %	0 %	5 %	12 %	24 %	**38** %	21 %
66+	0 %	1 %	3 %	1 %	5 %	14 %	**76** %

From the comparison result in Table 1, we can see that, a significant improvement over the baseline can be made about 8%. This result reveals that the local age modeling could be advantageous over the global age modeling in the presence of huge variation in face appearance (as shown in Fig. 3). In addition, it is demonstrated that our proposed method outperforms the other state-of-the-arts in Table 1.

Table 2 shows the confusion matrix for the proposed age classification. We observe that, for the age group '0-2' and '66+', the proposed method achieves acceptable classification accuracies (e.g., around 75%). On the other hand, recognizing the other age groups are relatively difficult as these age groups could be confused with both the younger and elder age groups.

4. CONCLUSION

This paper presented a new age classification method that aimed to deal with the large variation in each age group in realistic face images. To reduce the effect of the factors unrelated to age, only a few important local models (those had similar appearance to test face) of each age group affected the age classification. Experimental results using a real-world aging database verified the effectiveness of the proposed method using local age group modeling.

5. ACKNOWLEDGEMENT

This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2014.

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

6. REFERENCES

[1] Y. Fu, G. Guo, and T. S. Huang, “Age Synthesis and Estimation via Faces: A Survey,” *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 32, no. 11, pp. 1955-1976, 2010.

[2] Y. Kwon and N. Lobo, “Age Classification from Facial Images,” *Proc. IEEE Int’l Conf. on Computer Vision and Pattern Recognition,* 1994.

[3] Y. Kwon and N. Lobo, “Age Classification from Facial Images,” *Computer Vision and Image Understanding,* vol. 74, no. 1, pp. 1-21, 1999.

[4] A. Lanitis, C. Taylor, and T. Cootes, “Toward Automatic Simulation of Aging Effects on Face Images,” *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 24, no. 4, pp. 442-455, 2002.

[5] [Online]. Available: http://fipa.cs.kit.edu/befit/.

[6] T. Cootes, G. Edwards, and C. Taylor, “Active Appearance Models,” *Proc. European Conf. Computer Vision,* vol. 2, pp. 484-498, 1998.

[7] X. Gao, Y. Su, X. Li, and D. Tao, “A Review of Active Appearance Models,” *IEEE Trans. Sys. Man Cyber Part C,* vol. 40, no. 2, pp. 145-158, 2010.

[8] X. Geng, Z. -H. Zhou, and K. Smith-Miles, “Automatic Age Estimation Based on Facial Aging Patterns,” *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 29, no. 12, pp. 2234-2240, 2007.

[9] X. Geng, Z. -H. Zhou, and L Wang, “Facial Age Estimation by Nonlinear Aging Pattern Subspace,” *in Proc. ACM Conf. Multimedia,* 2008.

[10] K. Luu, K. Ricanek Jr, T. D. Bui, and C. Y. Suen, “Age estimation using Active Appearance Models and Support Vector Machine regression,” *IEEE Int. Conf. Biometrics: Theory, Applications, and Systems,* 2009.

[11] A. gallagher and T. Chen, “Understanding Images of Groups of People,” *IEEE Int’l Conf. on Computer Vision and Pattern Recognition,* 2009.

[12] [Online]. Available: http://www.flickr.com.

[13] C. Shan, “Learning Local Features for Age Estimation on Real-life Faces,” *in Proc. ACM Int’l Workshop. on Multimodal Pervasize Video Analysis,* 2010.

[14] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face Description with Local Binary Pattern: Application to Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 28, no. 12, pp. 2037-2041, 2006.

[15] L. Shen, L. Bai and M. C. Fairhurst, “Gabor Wavelets and General Discriminant Analysis for Face Identification and Verification,” *Image and Vision Computing,* vol. 25, no. 5, 2007.

[16] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” *IEEE Int’l Conf. on Computer Vision and Pattern Recognition,* 2001.

[17] J. Gong and D. W. Oard, “Selecting Hierarchical Clustering Cut Points for Web Person-name Disambiguation,” *in Proc. ACM Conf. Research and Development in Information Retrieval,* SIGIR’09. 2009.

[18] J. Ylioinas, A. Hadid, and M. Pietikäinen, “Age Classification in Unconstrained Conditions Using LBP Variants,” *IEEE Int’l Conf. on Pattern Recognition,* 2012.

[19] T. Bylander and L. Tate, “Using Validation Sets to Avoid Overfitting in AdaBoost,” *International Florida Artificial Intelligence Research Society Conference, AAAI Press,* 2006.

[20] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *In knowledge Discovery and Data Mining,* pp. 2, 1998.

[21] P. N. Belhumeur, J. P. Hesphanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 9, no. 7, pp. 711-720, 1997.

[22] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *J. Cognit. Neurosci.,* vol. 3, no. 1, pp. 71-86, 1991.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification.* Wiley, New York, 2000.

[24] C. S. Manning and H. Schütze, “Foundations of Statistical Natural Language Processing,” The MIT Press, pp. 500-512, 1999.

[25] M. Pawlak and M. Ng, “On Kernel and Radial Basis Function Techniques for Classification and Function Recovering,” *IEEE Int’l Conf. on Pattern Recognition,* 1994.

[26] B. Moghaddam, “Principal Manifolds and Probabilistic Subspaces for Visual Recognition,” *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 24, no. 6, pp. 780-788, 2002.

[27] F. Alnajar, C. Shan, T. Gevers, and J, Geusebroek, “Learning-Based Encoding with Soft Assignment for Age Estimation under Unconstrained Imaging Conditions,” *Imaging and Vision Computing,* vol. 30, no. 12, pp. 946-953, 2012.

[28] F. Gao and H. Ai, “Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method,” *Proc. Int’l Conf. Advanced in Biometrics,* pp. 132-141, 2009.

[29] J.-G. Wang, W.-Y. Yau, and H. L. Wang, “Age Categorization via ECOC with Fused Gabor and LBP Features,” *in Proc. IEEE Workshop Applic. Comput. Vis.,* 2009.