pp.2577-2581 Proceedings of ICASSP 2017

TWO-STAGE FACIAL AGE PREDICTION USING GROUP-SPECIFIC FEATURES

Jhony K. Pontes[⋆†] *Clinton Fookes*[⋆†] *Alceu S. Britto Jr.*[†] *Alessandro L. Koerich*[⋆]

[⋆†] Image and Video Research Laboratory, QUT, Australia, {*j.kaesemodel, c.fookes*}*@qut.edu.au*
[†] Postgraduate Program in Informatics, PUCPR, Brazil, *alceu@ppgia.pucpr.br*
[⋆] Dep. of Software and IT Engineering, ETS, Canada, *alessandro.koerich@etsmtl.ca*

ABSTRACT

A novel two-stage age prediction approach with group-specific features is proposed in this paper. Aging process is captured through a highly discriminating feature representation that models shape, appearance, skin spots, and wrinkles. The two-stage method consists of a multi-class Support Vector Machine (SVM) to predict the age bracket while the final age prediction is carried out using Support Vector Regression (SVR). The novelty of our work is that the feature extraction is group-specific and can therefore be tailored to each age bracket in the specific age prediction step. The FG-NET Aging dataset was used to evaluate the proposed method and an impressive mean absolute error (MAE) of 3.98 was achieved. Our approach outperforms the current state-of-the-art while increasing the robustness to blur, expression and lighting variation with local phase features.

Index Terms— Age prediction, Two-stage classifier, Group-specific features, Local phase features, Overlapping

1. INTRODUCTION

Age prediction from facial images is increasingly receiving attention to solve age-invariant person identification, age-based access control, age-adaptive targeted marketing, amongst other applications. Moreover, many complex diseases are associated with aging. Motivated by this, Chen *et al*. [1] found that facial morphology features are correlated with health indicators in the blood. Facial features therefore are more reliable aging bio-markers than blood profiles and can better reflect general health status than chronological age. This is a potential application for age prediction that may help to improve our health. Facial aging is a complex process and it is generally slow and irreversible [2] since different factors can influence it such as gender, heredity, ethnicity, lifestyle, environment, etc. All these factors and also different perturbations on images, such as expression, lighting, occlusion, pose, and blur, make determining a subjects's age from their face very challenging.

Several age prediction approaches have been proposed in recent years. For instance, Lanitis *et al.* [3] proposed the Weighted Appearance Specific (WAS) which represents the aging pattern by a quadratic function. Geng *et al.* [4] proposed the AGES algorithm which uses a sequence of individual's facial images for the aging modeling. Duong *et al.* [5] presented a hierarchical approach that combines SVM and SVR to improve the performance of age prediction. Choi *et al.* [6] proposed a hierarchical age prediction to handle hard boundaries between age brackets. They also exploited Local Binary Patterns (LBP) and Gabor wavelets for the extraction of appearance features. Recently, Liu *et al.* [7] proposed a hybrid constraint SVR that uses fuzzy age labels in combination with the real ones to train it. Deep learning aging representation approaches were presented by Wang *et al.* [8] and Chang and Chen [9]. The last proposed a cost-sensitive ordinal hyperplanes ranking (CSOHR) algorithm for age prediction and also the scattering transform (ST) for feature representation that is based on convolutional neural networks (CNN). Dibeklioglu *et al.* [10] combined facial dynamics derived from facial expressions and combined it with the appearance features to train the classifiers and regressors. Pontes *et al.* [11] proposed a flexible hierarchical approach to deal with the hard boundaries between the age brackets and investigated multiple features to represent the facial aging.

This paper builds upon the Pontes *et al.*'s work [11] which presented an age prediction approach based on age bracket classification followed by detailed age prediction. The two-stage approach first classifies a subject into a determined age bracket and then a specific regressor is selected to estimate its age. As in [11], we perform a varying overlapping of age ranges in the specific age prediction step to tackle the problem of hard boundaries between age brackets imposed by the two-stage method. An individual that is misclassified during the classification step can still have its age well predicted by a specific regressor due to the varying overlapped age range [11]. The novelty of our work is that we perform the training of the regressors in the specific age prediction step using different feature sets. Our hypothesis is that group-specific features can capture better the particularities of a specific age range than general features and that might lead to better results on age prediction. In the proposed approach, we use global features to encode the craniofacial growth using the AAM method since it provides appearance and shape information. Holistic AAM features however do not contain

enough skin texture information and to overcome this, a local approach is used to cope with it. Skin aging is the most perceptible change from adulthood to old age, and to deal with wrinkles and skin spots the LBP, the Gabor wavelets and the LPQ techniques are used to extract local features. Finally, global and local features are combined to form a hybrid feature vector to have a more discriminating facial aging representation. We demonstrate the superior performance and robustness of our assumption on the FG-NET Aging dataset.

This paper is organized as follows. Section 2 presents the proposed approach. The results are presented in Section 3. Finally, the conclusions are stated in the last section.

2. PROPOSED APPROACH

An overview of the proposed method is shown in Fig. 1. Images from the FG-NET Aging dataset were not taken in controlled conditions. For that reason, every image is converted to grayscale to decrease the influence of inconsistent colors. The dataset provides 68 landmarks for each image that include the center of the eyes which is used to apply a non-reflective similarity transformation to normalize the images. The facial images are scaled to the same inter-pupillary distance [11] and then a mask is applied to delimit and crop the skin regions of interest. An important component to describe wrinkles and spots is the local features. The same eleven skin regions highly correlated with aging proposed by [11] were selected to capture the aging process (*i.e.* corners of the eyes and the mouth, forehead, cheeks, chin, and nose). Fig. 2 shows an example of the eleven skin regions cropped.

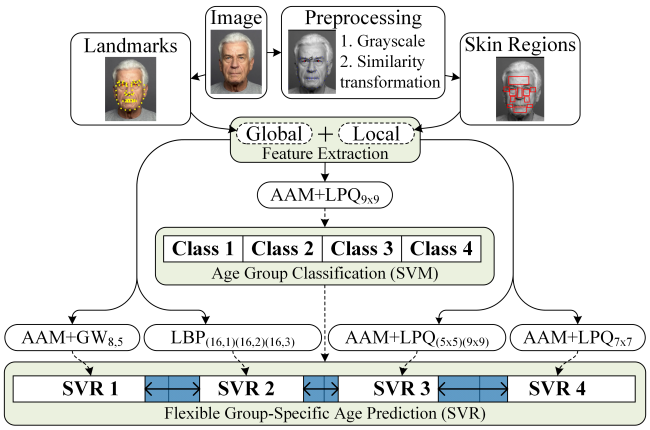
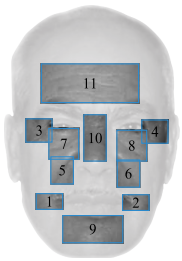


Fig. 1. Overview of the proposed age prediction method.

Active Appearance Model is a widely used statistical method for facial modeling and feature extraction, where the shape and texture variability are captured from a representative training set. Principal component analysis (PCA) on shape and texture allows to produce a parametric model that describes the global features of a face. Thus, to capture shape and appearance features related to aging, the AAM is



1. Left corner of the mouth lines (36x21)
2. Right corner of the mouth lines (36x21)
3. Left periorbital lines (36x31)
4. Right periorbital lines (36x31)
5. Left nasolabial folds (31x36)
6. Right nasolabial folds (31x36)
7. Left cheek lines (41x41)
8. Right cheek lines (41x41)
9. Chin crease (81x36)
10. Top nose (31x61)
11. Horizontal forehead lines (131x51)

Fig. 2. Example of the eleven skin regions and its definition.

employed as global features in the proposed method. Local features are extracted by the LBP, the Gabor wavelets and the LPQ operators. LPQ is a powerful feature extractor that has been used as an alternative to the widely used LBP [12], mainly due to its robustness not only to blur but also to facial expression and lighting changes present in real-world images. Consequently, the LPQ descriptor is employed in this work to capture wrinkles through local phase features.

In LPQ the phase is examined in local M-by-M neighbourhoods N_x at each pixel position x of the image $f(x)$. These local spectra are computed using a short-term Fourier transform. The phase information in the Fourier coefficients is captured by observing the signs of the real and imaginary parts of each component by using a scalar quantizer $q_j(x)$ that assumes 1 in case of $g_j(x) \geq 0$, otherwise it assumes 0, where $g_j(x)$ is the jth component of the vector, $G_x = [Re\{F_x\}, Im\{F_x\}]$. The binary coefficients $q_j(x)$ are represented as integer values (0-255) using the coding defined by the equation $f_{LPQ}(x) = \sum_{j=1}^{8} q_j(x)2^{j-1}$. As a result, we have the label image f_{LPQ} whose values are the blur invariant LPQ labels. Skin details such as thin wrinkles and spots are captured by the LBP that can detect microstructures [13]. It labels the pixels of an image by thresholding the neighbourhood of each pixel and it considers the result as a binary number. Since LBP considers uniform binary patterns, it is robust to lighting changes and noise which makes the descriptor very practical for real applications. A set of Gabor wavelets are employed to capture local features related to wrinkles due to its robustness to noise such as hair, beard, shadows, etc. [11, 14, 15]. The Gabor transformation at a particular image position is calculated through a convolution with the wavelets, and the local features derive from the magnitude of the resulting complex image.

All the local descriptors are applied in every cropped and normalized skin region, and once the local features of every region have been extracted independently, the feature vectors are concatenated into a single vector. In addition, several combinations between different techniques are performed in an attempt to capture unique advantages of each descriptor. Feature-level fusion has been successfully applied in biometrics [16], images [17], facial expressions [18], etc., and it is

also employed in this work. Since the local feature vector extracted by the Gabor wavelets, the LBP and the LPQ is high dimensional, only the principal components are chosen with the PCA. The z-score is then applied to normalize the lower dimensionality features [16]. The feature fusion is built up through the concatenation of the normalized feature vectors.

Aging differs according to the age brackets of a person. Wrinkles are usually found in old subjects, while geometric features normally change during childhood and adulthood. Two-stage age prediction based on age-bracket specific classifiers has shown improved results and better ability to deal with these age-related facial features [6, 11]. Motivated by this, we argue that group-specific features in the specific age prediction can better represent age brackets and it can yield better results. As in [11], we first classify a facial image into a determined age class by using a SVM and then a SVR is used to estimate the specific age. The novelty is that the features are group-specific so it can be different for each SVR class in the specific age prediction step. Different amounts of age range overlapping are considered for the specific age prediction step to decrease misclassification caused by the classification step. The overlapped classes are defined according to the misclassification errors and the amount of overlapping can vary between each pair of SVRs.

3. EXPERIMENTAL RESULTS

The proposed method was evaluated on the FG-NET dataset [19] which contains 1,002 facial images of 82 subjects from different races with ages between 0 to 69 years old. The face images were captured with variation of facial expression, pose, and lighting, as well as the presence of blurring and occlusion. All images have 68 manually annotated landmark points characterizing the shape features of the subjects. The dataset was uniformly split into 75% for training and 25% for test to perform the experiments. The SVM and the SVRs were set to use a linear kernel and their parameters were found by a grid search using 5-fold cross validation. Once the dataset has a limited amount of data, the Leave-One-Person-Out (LOPO) cross-validation protocol was also used to evaluate the performance of the proposed method. This guarantees that a subject is not in the training and the test set simultaneously, so the classifier does not learn individual characteristics, thus decreasing the dependence on data on the experimental results.

For the age group classification step we defined four age brackets based on the data distribution with the aim of having enough data for learning and evaluation. This also allows to have a group for children, young adults, adults, and seniors, especially useful for the FG-NET dataset since it has a wide age range. The classes were set as: **Class 1**: 0-13 (total of 513 images); **Class 2**: 14-21 (233 images); **Class 3**: 22-39 (187 images); and **Class 4**: 40-69 (69 images). The age ranges of the SVRs were defined based on the same ranges described previously but with different amounts of overlapping depending on the age ranges. The amounts of overlapping between the SVRs were defined experimentally on the training dataset.

The AAM was generated using the 68 landmarks provided for each image. A multi-resolution model based on a Gaussian pyramid was built to deal with the resolution changes present in the dataset [20]. Some relevant age features are not sufficiently reconstructed such as the wrinkles in the corner of the eyes and the nasolabial lines due to the loss of information during the PCA. A local feature descriptor is further necessary to have a better facial aging representation since the representation power of the AAM is limited. Gabor wavelets were evaluated in this work to detect lines, and since a face can contain wrinkles with different thicknesses and directions it is necessary to generate wavelets with different scale factors and orientations. Local features of the eleven skin regions were also extracted with the $\text{LBP}^{u2}_{P,R}$ operator, where P and R are the circular neighborhood parameters, and P can assume eight or sixteen pixels while R can vary from one to three pixels. The naming $u2$ means that only uniform patterns are used to decrease noise in the local texture structures. The LBP descriptor is very discriminating and captures mostly edges and spots information on the skin if the image is blur-free. In the experiments using the LPQ the features were extracted and evaluated using different window sizes (M). The LPQ is locally computed for a window in each skin region position providing an histogram of 256 codes. The code values are generally higher for young subjects since the phase orientations are similar due to a smoother skin texture [11].

The performance of the proposed approach was evaluated using the MAE and the cumulative score (CS) [21] metrics. MAE is the average of the absolute errors between the estimated age and the ground-truth age. The CS allows the comparison of performance at different absolute error levels and it is defined as $CS(l) = (N_{e \leq l}/N) \times 100\%$, where $N_{e \leq l}$ is the number of test images on which the age prediction makes an absolute error lower or equal to l years. Several feature sets combinations were evaluated, either global and local features alone or the combination of both to have a better facial representation. The highest accuracy, 68.80%, was achieved with the global and local features AAM+$\text{LPQ}_{9\times9}$. In order to recover the errors caused by the age bracket classification step, varying overlapped age ranges in the specific age prediction step are used in this work. The classification errors were analyzed to decide on the initial amount of overlapping, from which, through the increasing or decreasing of such amount, the values were chosen based on the best results in terms of MAE of the age prediction method.

Fig. 3 shows the performance in terms of MAE of the specific age prediction with varying overlapped regions between the age brackets on various feature sets. Looking individually at each SVR, the best SVR 1 achieved a MAE of 3.41 years with the AAM+$\text{GW}_{8,5}$ feature set. The best SVR 2 achieved a MAE of 4.64 years with the $\text{LBP}_{(16,1)(16,2)(16,3)}$. The best SVR 3 achieved a MAE of 6.63 years with the

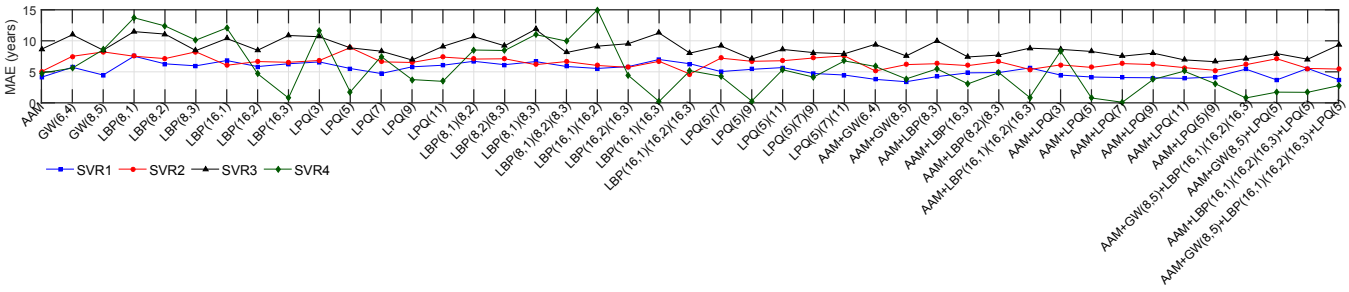


Fig. 3. Performance in different feature sets of the specific age prediction method with varying overlapped regions.

AAM+LPQ$_{(5\times5)(9\times9)}$ combination. Finally, the best SVR 4 achieved a MAE of 0.09 years with the AAM+LPQ$_{7\times7}$ feature set. It is possible to notice from Fig. 3 that the SVR 1 has the best overall MAE since most of the images from the training set are into its age brackets. In the other hand, the MAE of the SVR 4 widely oscillates across the extremes due to the limitation of data for training the 40-69 years old class. Fig. 3 clearly shows that the MAE increases with age as expected, since the main features associated with face shape greatly changes from 0 to 18 years old with no significant changes in people older than 18 years. These shape features make estimating a subjects's age easier than looking at facial skin features that only change significantly in older people.

Our group-specific features approach for the specific age prediction step can now be performed since we know the results of each SVR in several feature sets. Instead of training all the SVRs with the same features as in other works, we train each SVR with an independent feature set according to its performance in a specific age bracket. Performing the two-stage age prediction using the best SVRs as described earlier we achieved an impressive MAE of only 3.98 years. This result cleary shows that using group-specific features in the specific age prediction step can improve the final performance of a two-stage age prediction.

Table 1. Comparison of the proposed method to the previous works reported on the FG-NET Aging dataset.

Method	MAE	Method	MAE
WAS [3]	8.06	Choi *et al.* [6]	4.65
AGES [4]	6.77	PLO [22]	4.82
RUN [23]	5.78	CA-SVR [24]	4.67
Ranking [23]	5.33	Han *et al.* [25]	5.10
LARR [26]	5.07	HC-SVR [7]	5.28
SVR [26]	5.66	Wang *et al.* (CNN) [8]	4.26
MTWGP [27]	4.83	ST+CSOHR (CNN) [9]	4.70
OHRank [28]	4.85	Pontes *et al.* [11]	4.50
Duong *et al.* [5]	4.74	**Group-specific features**	**3.98**

The MAE for the proposed approach and other methods reported on the FG-NET dataset are given in Tab. 1. It can be noticed that the proposed approach has the lowest MAE when compared with other state-of-the-art methods. Deep learning techniques for age estimation based on CNNs have been investigated by Wang et al. [8] and Chang and Chen [9]. Nevertheless, our method still outperform their approaches by a margin. Fig. 4 shows the CS at error levels from 0 to 15 years. Note that it moved up by using a group-specific features in the specific age prediction step. We have also evaluated the proposed strategy on the MORPH Album 2 dataset used in [11] and we have observed a reduction of $\sim$5% in the MAE. This demonstrates that the proposed approach is robust to other datasets. To the best of our knowledge, this is the first work using group-specific features in the specific age prediction step yielding promising results.

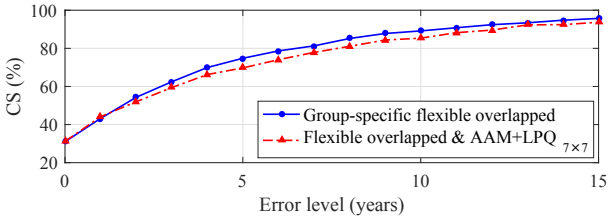


Fig. 4. Cumulative scores of the proposed approach (blue) and the two-stage age prediction proposed in [11] (red) at error levels from 0 to 15 years.

4. CONCLUSIONS

This paper has presented a novel two-stage age prediction approach that uses group-specific features during the specific age prediction step. Feature extraction can therefore be tailored to each age group meaning that the type of features used can be different for each SVR class. Experimental results on the FG-NET dataset show that our method outperforms other state-of-the-art systems for age prediction while increasing the robustness to blur, expression and lighting variation with local phase features. Future work will address techniques based on deep learning to enhance the age prediction.

5. REFERENCES

[1] W. Chen et al., "Three-dimensional human facial morphologies as robust aging markers," *Cell Research*, pp. 1–14, 2015.

[2] A. Stone, "The Aging Process of the Face & Techniques of Rejuvenation," http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtm, 2010.

[3] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE TPAMI*, vol. 24, pp. 442–455, 2002.

[4] X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE TPAMI*, vol. 29, pp. 2234–2240, 2007.

[5] C. N. Duong, K. G. Quach, K. Luu, H. B. Le, and K. Ricanek, "Fine tuning age estimation with global and local facial features," in *in IEEE ICASSP*, 2011, pp. 2032–2035.

[6] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, pp. 1262–1281, 2011.

[7] J. Liu, Y. Ma, L. Duan, F. Wang, and Y. Liu, "Hybrid constraint SVR for facial age estimation," *Signal Processing*, vol. 94, pp. 576–582, 2014.

[8] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," *in IEEE WACV*, pp. 534–541, 2015.

[9] K-Y. Chang, , and C-S. Chen, "A learning framework for age rank estimation based on face images with scattering transform," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 785–798, 2015.

[10] H. Dibeklioglu, F. Alnajar, A. A. Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1928–1943, 2015.

[11] J. K. Pontes, A. S. Britto Jr., C. Fookes, and A. L. Koerich, "A flexible hierarchical approach for facial age estimation based on multiple features," *Pattern Recognition*, vol. 54, pp. 34–51, 2016.

[12] T. Ahonen and E. Rahtu, "Recognition of blurred faces using local phase quantization," *in IEEE CVPR*, pp. 1–4, 2008.

[13] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, pp. 971–987, 2002.

[14] K. B. Vinay and B. S. Shreyas, "Face recognition using Gabor wavelets," *Signals, Systems and Computers*, pp. 593–597, 2006.

[15] M. Bereta, P. Karczmarek, W. Pedrycz, and M. Reformat, "Local descriptors in application to the aging problem in face recognition," *Pattern Recognition*, vol. 46, pp. 2634–2646, 2013.

[16] A. Ross and R. Govindarajan, "Feature level fusion in biometric systems," in *in Proc. of the Biometric Cons. Conf.*, 2004.

[17] A. Marini, A. J. Turatti, A. S. Britto Jr., and A. L. Koerich, "Visual and acoustic identification of bird species," *in IEEE ICASSP*, pp. 2309–2313, April 2015.

[18] T. H. H. Zavaschi, A. S. Britto Jr., L. E. S. Oliveira, and A. L. Koerich, "Fusion of feature sets and classifiers for facial expression recognition," *Expert Systems with Applications*, vol. 40, no. 2, pp. 646–655, 2013.

[19] "The FG-NET aging database," http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html, 2010.

[20] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE TPAMI*, vol. 23, pp. 681–685, 2001.

[21] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE TPAMI*, vol. 32, pp. 1955–1976, 2010.

[22] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," *In CVPR*, p. 2570–2577, 2012.

[23] S. Yan, H. Wang, T. Huang, Q. Yang, and X. T. Xiaoou, "Ranking with uncertain labels," in *in IEEE ICME*, 2007, pp. 96–99.

[24] K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," *In CVPR*, p. 2467–2474, 2013.

[25] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *in IEEE ICB*, 2013.

[26] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process*, p. 1178–1188, 2008.

[27] Y. Zhang and D.-Y. Yeung, "Multi-task warped gaussian process for personalized age estimation," *In CVPR*, p. 2622–2629, 2010.

[28] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," *In CVPR*, p. 585–592, 2011.