09-16

15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Multi-layer Age Regression for Face Age Estimation

Choon-Ching Ng[a], Yi-Tseng Cheng[b], Gee-Sern Hsu[b], Moi Hoon Yap[c]
[a]Panasonic R&D Center Singapore, Singapore
[b]National Taiwan University of Science and Technology, Taipei, Taiwan
[c]Manchester Metropolitan University, Manchester M1 5GD
[a]choonching.ng@sg.panasonic.com, [b]jison@mail.ntust.edu.tw, [c]m.yap@mmu.ac.uk

Abstract

Face features convey many personal information that promote and regulate our social linkages. Age prediction using single layer estimation such as aging subspace or a hybrid pattern is limited due to the complexity of human faces. In this work, we propose Multi-layer Age Regression (MAR) where the face age is predicted based on a coarse-to-fine estimation using global and local features. In the first layer, Support Vector Regression (SVR) performs a between group prediction by the parameters of Facial Appearance Model (FAM). In the second layer, a within group estimation is performed using FAM, Bio-Inspired Features (BIF), Kernel-based Local Binary Patterns (KLBP) and Multi-scale Wrinkle Patterns (MWP). The performance of MAR is assessed on four benchmark datasets: FGNET, MORPH, FERET and PAL. Results showed that MAR outperforms the state of the art on FERET with a Mean Absolute Error (MAE) of 3.00 (±4.14).

1   Introduction

As humans age, myriad changes occur chronically within the craniofacial complex [1]. Notable soft tissue modifications may be seen across each decade of adult life that passes. As well, subtle hard tissue or bony changes slightly alter an overall shape of the human face, mainly in the dentoalveolar region (portion of the alveolar bone immediately about teeth). These age-related changes affect the accuracy and efficacy of face-related applications [2, 3]. The real-world applications are very rich and attractive, existing facts and attitudes from the perception field reveal the difficulties and challenges of automatic age synthesis and estimation by computer [4].

The process of age estimation attempts to label a face image automatically with the exact age (year) or the age group (year range) of the individual face. By deriving significant features from faces of known ages, the age of an individual face can be estimated by solving the inverse problem using the same feature-extraction technique. Although many algorithms have been proposed since 1994 [5], age estimation is still a challenging problem due to three reasons [6]. First, facial age progression is uncontrolled and personalised. Characteristics of aging variation cannot be captured easily due to the large variations conveyed by human faces. Facial aging effects are manifested in different forms during different ages. While facial aging effects are predominantly manifested in the form of facial shape variation during formative years, textural variations in the form of wrinkles and other skin artefacts take precedence over shape variations during later stages of adulthood. Hence, facial aging can be described as a problem of characterising facial shape and facial texture as functions of time. Second, there is no complete facial aging dataset with chronological ages. Developing facial growth models or building characterisations of facial aging begins with identifying the appropriate form of data that provide a fair description of the event. The data could be individual-specific or population-specific. It is hard to collect a large facial image set of people throughout their life which are sufficient to present detailed aging progression. Third, it is difficult to define an absolute aging pattern which can be used to quantify one particular age.

This paper proposes multi-layer age regression for face age estimation. The performance is compared with the state-of-the-art algorithms and assessed on four popular datasets: FGNET [7], FERET [8], MORPH [9] and PAL [10].

2   Feature Representation

This section presents four popular state-of-the-arts feature descriptors for face age estimation: Facial Appearance Model (FAM) [11], BIF [12], Kernel-based Local Binary Patterns (KLBP) [13] and Multi-scale Wrinkle Patterns (MWP) [14].

2.1   Facial Appearance Model

FAM is a generative parametric model that consists of shape, texture and combined appearance of a human face. It is a model where PCA is used to project high dimension of face shapes and textures into a low dimension of principal component parameters. The pertinent equations of FAM are summarized, for full description about this method, please refer to Cootes et al. [11]. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesized shape and texture of a face image in the reference frame, and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the principal component scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s \qquad (1)$$

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \mathbf{b}_t \qquad (2)$$

where Φ_s and Φ_t are matrices of column eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \left[\begin{array}{c} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{array} \right] = \left[\begin{array}{c} \mathbf{W}_s \Phi_s^T (\mathbf{s} - \bar{\mathbf{s}}) \\ \Phi_t^T (\mathbf{t} - \bar{\mathbf{t}}) \end{array} \right] \qquad (3)$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$. To make the normalised measures of pixel distance and pixel intensities commensurate, the shape model scores are typically weighted by the square root of the ratio between the sums of the texture and shape eigenvalues.

To recover any correlation between shape and texture, the two eigen-spaces are usually coupled through a third principal component transform as

$$\mathbf{b} = \Phi_c \mathbf{c} = \left[\begin{array}{c} \Phi_{c,s} \\ \Phi_{c,t} \end{array} \right] \mathbf{c} \qquad (4)$$

and $\mathbf{b}$ is the FAM features of each image.

2.2 Bio-inspired Features

BIF have shown good performance for age estimation [12]. A specially-designed BIF with two layers: the simple layer S_1 and complex layer C_1. The S_1 units correspond to the cells in the primary visual cortex. They are typically implemented with the convolution of an image with a Gabor filter defined as

$$G\left(x, y\right) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\zeta^2}\right) \times \cos\left(\frac{2\pi}{\chi} x'\right) \qquad (5)$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$ are the rotations of the Gabor filters for angle θ which varies between 0 and π. The aspect ratio is fixed as $\gamma = 0.3$, the effective width ζ, the wavelength χ as well as the filter sizes s were adjusted as in [12]. The orientation θ varies from 0 to π uniformly with different intervals, resulting in different numbers of total orientations, such as 4, 6, 8, 10, and 12. The C_1 units correspond to cells which are robust to shift and scale variations. They can be calculated by pooling over the preceding S_1 units with the same orientation but at two successive scales. "MAX" pooling operator and "STD" normalization operator are two C_1 features extracted from the S_1 layer. Both operators in the C_1 layer are finally concatenated into a single feature vector $\mathbf{f}$ as,

$$\mathbf{f} = \Upsilon_{\text{MAX,STD}}\left(\text{Re}\left(I * G\right)\right) \qquad (6)$$

where I is a 2D image and Υ denotes two consecutive operations of "MAX" and "STD" on the convolved image. In this work, the feature vector of $\mathbf{f}$ is 7464 units.

2.3 Kernel-based Local Binary Patterns

According to Ylioinas et al. [13], the sparse nature of LBP representation is improved by the proposed kernel estimator. A face image is first divided into a set of LBP overlapping patches of a size 13×13 pixels, each patch overlapping its vertical and horizontal neighbours by 4 units. With a face image of size 76×76, this results 64 patches [13]. In a particular patch, it is decomposed into two complementary components, sign and magnitude. The sign component is coded using the conventional LBP operator defined as

$$\text{LBP}_{P,R}^S = \sum_{p=0}^{P-1} \mathcal{T}\left(g_p - g_c\right) 2^p \qquad (7)$$

where g_c corresponds to the grey value of the center pixel, g_p refers to grey values of P equally spaced pixels on a circle of radius R, and $\mathcal{T}$ defines a thresholding function with $\mathcal{T}(x) = 1$ if $x \geq 0$ and $\mathcal{T}(x) = 0$ otherwise. The magnitude component is defined as

$$\text{LBP}_{P,R}^M = \sum_{p=0}^{P-1} \mathcal{T}\left(\hat{m}_p, \hat{c}\right) 2^p \qquad (8)$$

where $\hat{m}_p$ is the magnitude of local pixel difference and $\hat{c}$ a predetermined threshold value of LBP usually set as the mean value of local pixel differences in the whole image. As the magnitude operator encodes the difference in local pixel intensities, it gives a measure of contrast. The key idea of LBP is to gain more comprehensive image representation by combining these two complementary descriptions.

In order to estimate the probability distribution of LBP-like random variables, the kernel is defined as

$$K_{\hat{h}}\left(\hat{l}|\hat{l}'\right) = \hat{h}^{P - \tilde{d}\left(\hat{l}, \hat{l}'\right)} \left(1 - \hat{h}\right)^{\tilde{d}\left(\hat{l}, \hat{l}'\right)} \qquad (9)$$

where $\hat{l}$ and $\hat{l}'$ are both P-dimensional binary variables, $\tilde{d}$ is the Hamming distance between them, and $\hat{h}$ is a bandwidth parameter. Finally, using the given kernel, KLBP pattern is defined as

$$\mathbf{g} = \{\Psi_1 K_{\hat{h}}, \Psi_2 K_{\hat{h}} \cdots, \Psi_n K_{\hat{h}}\} \qquad (10)$$

where $\Psi \in \{\text{LBP}_{P,R}^S, \text{LBP}_{P,R}^M\}$ and n is the total number of KLBP features of each image.

2.4 Multi-scale Wrinkle Patterns

Ng et al. [14] proposed MWP features for face age estimation. By deriving wrinkles with the multi-scale filters across ten face regions, wrinkle patterns are generated. First, they locate facial landmarks by using the Face++ detector and then normalize the face by using a linear transformation. A face template which consists of ten predefined wrinkle regions is defined. Then, for each region, they detect wrinkles using Hessian filter [16] and cross-sectional profile, $\mathcal{V}$, which is defined as

$$\mathcal{V}_{\hat{s}}\left(\theta\right) = I_{\hat{s}}\left(p_1\right) + I_{\hat{s}}\left(p_2\right) - I_c - I_{\hat{s}}\left(\theta\right) \qquad (11)$$

where I_c is the current pixel, $I_{\hat{s}}\left(\theta\right)$ is the candidate pixel in a particular direction θ, $I_{\hat{s}}\left(p_1\right)$ and $I_{\hat{s}}\left(p_2\right)$ are background pixels away from the candidate pixel, and $\hat{s}$ is the filter scale. If the I_c and $I_{\hat{s}}\left(\theta\right)$ belong to the set of wrinkles, $\mathcal{V}$ has a large positive value. If the I_c and $I_{\hat{s}}\left(\theta\right)$ belong to the background, they have similar values, thus $\mathcal{V}$ has a negative value or is near to zero. For each $\hat{s}$, confidence array of each pixel is increased by one if $\mathcal{V}$ is positive and larger than a predefined threshold. Wrinkles are detected by selecting the maximum confidence array of each pixel. Finally, the detected wrinkles of each region are represented as wrinkles intensity, $\mathfrak{f}_i$, and density, $\mathfrak{g}_i$, that defined as

$$\mathfrak{f}_i = \log \sum_{x=1}^{wt} \sum_{y=1}^{ht} I_i\left(x, y\right) \qquad (12)$$

$$\mathfrak{g}_i = \frac{area_1\left(i\right)}{area_2\left(i\right)} \qquad (13)$$
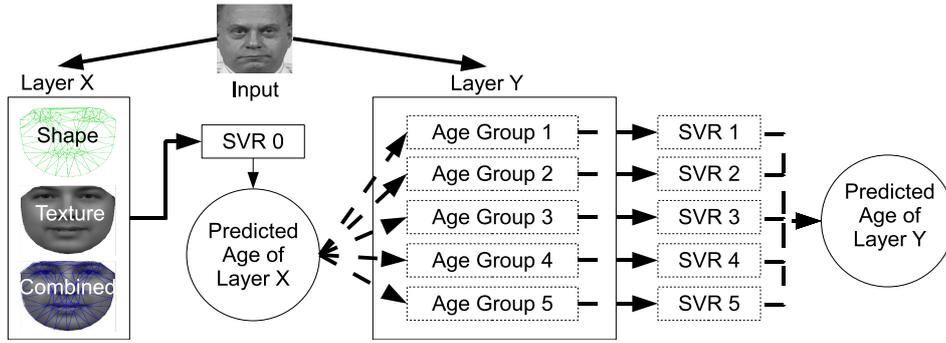
Figure 1. Flow chart of MAR. It consists of two layers: layer X and layer Y. Layer X is the parameters of FAM while layer Y is the features of FAM, BIF, KLBP and MWP. First, SVR 0 estimates the age of input using FAM parameters. Then, the predicted age of layer X is used to select an age group of layer Y. Each age group has a fixed range of ages. If age group 3 is selected , then SVR 3 is used to train the features in age group 3 and finally predict the age of input in layer Y. The output of both layers is a real number (predicted age) instead of a class. Dotted arrow indicates only one of the age groups will be selected. Note that this figure was redrawn from FERET.

where wt and ht are the width and height of I, $area_1$ is the wrinkle area found in a particular region i and $area_2$ is the area of region i. MWP pattern is generated by combining both features into a single vector as

$$\mathbf{h} = \{\mathfrak{f}_1, \mathfrak{g}_1, \cdots, \mathfrak{f}_n, \mathfrak{g}_n, \} \tag{14}$$

where n is the total number of MWP features.

3 Multi-layer Age Regression

MAR consists of two layers, layer X and Y as shown in Figure 1. In layer X, features are represented by the FAM parameters, while in layer Y, features are represented by descriptors such as FAM, BIF, KLBP and MWP. Let Λ_X as the predicted age of layer X and the prediction is defined as

$$\Lambda_X = f(\mathbf{b}) \tag{15}$$

where $f(\cdot)$ is an estimation function by Support Vector Regression (SVR).

In the layer Y, a between-group classification is implemented based on Λ_X. Let the sorted features in different age groups as $\{\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_n\}$ where n is the total number of age groups of each dataset and Λ_Y as the predicted age of layer Y. The prediction is defined as

$$\Lambda_Y = \begin{cases} f_1(\mathbf{G}_1, \mathbf{d}) & \text{if } \Lambda_X < a_1 \\ f_2(\mathbf{G}_2, \mathbf{d}) & \text{if } a_1 \leq \Lambda_X < a_2 \\ f_3(\mathbf{G}_3, \mathbf{d}) & \text{if } a_2 \leq \Lambda_X < a_3 \\ f_4(\mathbf{G}_4, \mathbf{d}) & \text{if } a_3 \leq \Lambda_X < a_4 \\ f_5(\mathbf{G}_5, \mathbf{d}) & \text{if } a_4 \leq \Lambda_X \end{cases} \tag{16}$$

where f_i is one particular SVR model, $\mathbf{G}_i$ is the training set, $\mathbf{d}$ is the testing set of an input image, and both $\mathbf{G}_i$ and $\mathbf{d}$ are the same feature selected from $\{\mathbf{b}|\mathbf{f}|\mathbf{g}|\mathbf{h}\}$.

In Figure 1, SVR 0 estimates the age of input using FAM parameters. Then, the predicted age of layer X is used to select an age group of layer Y. Each age group has a fixed range. If age group 3 is selected, then SVR 3 is used to train the features in age group 3 and finally predict the exact age of I in layer Y. Based on such

design, the age of a face image is estimated by a coarse-to-fine approach where the first layer focuses on using regression to predict the age, which will then be used in selecting an age group, while layer Y concentrates on predicting the exact age within the group.

4 Experimental Results

In this work, four datasets, FGNET (with 1002 images), MORPH (with 2000 randomly selected images), FERET (with 2366 images) and PAL (with 576 images), are utilized for performance validation. For FGNET, a Leave One Person Out (LOPO) and 68 landmark points are used as the evaluation protocol while the remaining datasets are based on 10-folds cross validation and 83 landmark points of FACE++ detector [15]. Due to many works of FGNET are using 68 points in the experimental setup, the same landmark is used for a fair comparison. Both training and testing datasets are disjoint. In FAM modelling, a 95% of the parameters variability are preserved during the dimension reduction of PCA in the training set. FAM fitting is not considered in this experiment as the fitting error might increase the error of age estimation [17]. Due to different features have different input dimensions, the SVR parameters are varied across distinct descriptors and datasets. These parameters are derived from a grid search approach [13].

Table 1 shows a comprehensive analysis of face age estimation with and without MAR. Overall, the proposed MAR methods, FAM-FAM and FAM-BIF perform the best on FERET, with MAE of 3.00 ($\pm$4.14) and on PAL with the MAE of 3.43 ($\pm$2.71), respectively. However, FAM and BIF perform the best on FGNET and MORPH with MAE of 5.39 ($\pm$5.63) and 3.98 ($\pm$3.20) respectively.

The lowest MAE of each method is obtained on FERET. Overall, results showed that MAR improves MAE significantly on FERET compared to the existing methods. Although MAE of FAM-FAM is the lowest, other results are comparable. For example, the results of FERET showed that FAM-FAM hits a MAE of 3.00 compared to FAM-MWP is 3.36, FAM-BIF is 3.28 and FAM-KLBP is 3.29; the results of MORPH showed

Table 1. Experimental results of face age estimation with MAR. Bold - the lowest MAE within one particular dataset, italic - the lowest MAE of each method amongst the datasets.

Datasets	Single-layer Age Estimation, MAE (STD)				Multi-layer Age Estimation, MAE (STD)			
	FAM	BIF	KLBP	MWP	FAM-FAM	FAM-BIF	FAM-KLBP	FAM-MWP
FGNET	**5.39 (5.63)**	5.59 (5.97)	6.09 (6.43)	7.34 (7.54)	5.48 (6.67)	5.49 (6.61)	5.79 (6.67)	6.19 (6.85)
FERET	*3.34 (3.26)*	*3.57 (3.26)*	*3.91 (3.24)*	*4.16 (3.83)*	***3.00 (4.14)***	*3.28 (4.02)*	*3.29 (4.07)*	*3.36 (4.14)*
MORPH	3.99 (3.28)	**3.98 (3.20)**	4.02 (3.22)	5.16 (4.35)	4.18 (3.63)	4.06 (3.55)	4.06 (3.47)	4.08 (3.51)
PAL	6.96 (5.92)	5.94 (4.60)	6.09 (5.09)	7.65 (6.61)	6.72 (6.64)	**3.43 (2.71)**	6.33 (6.28)	6.63 (6.65)

that FAM-MWP achieves a MAE of 4.08 compared to FAM-FAM is 4.18, FAM-BIF is 4.06 and FAM-KLBP is 4.06. These results showed that MAR is comparable to the state of the art for face age estimation.

5 Conclusion

This paper has proposed a novel method, MAR, for face age estimation. It makes use of facial appearance model and local features in a coarse-to-fine estimation for face age estimation. Experiment results showed that MAR outperforms the state of the art with a MAE as low as 3.00 ($\pm$4.14) on FERET, and 3.43 ($\pm$2.71) on PAL. FAM performs the best on FGNET with MAE of 5.39 ($\pm$5.63) and BIF performs the best on MORPH with MAE of 3.98 ($\pm$3.20). Overall, the wrinkle based algorithm, MWP has the best computational time. Future works involve investigation on alternative descriptors, age group overlapping issue and increase the sample size for the computer algorithms.

Acknowledgment

References

[1] M. Zimbler, M. Kokoska, J. Thomas: "Anatomy and pathophysiology of facial aging," *Facial plastic surgery clinics of North America*, vol.9, no.2, pp.17987, 2001.

[2] A. Albert, K. Ricanek, E. Patterson: "A review of the literature on the aging adult skull and face: implications for forensic science research and applications," *Forensic Science International*, vol.172, no.1, pp.1-9, 2007.

[3] J. P. Farkas, J. E. Pessa, B. Hubbard, R. J. Rohrich: "The science and theory behind facial aging," *Plastic and Reconstructive Surgery Global Open*, vol.1, no.1.

[4] G. Guo, Y. Fu, C. Dyer, T. Huang: "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. on Image Processing*, 17(7):1178-1188, 2008.

[5] Y. Kwon, N. da Vitoria Lobo: "Age classification from facial images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.762-767, 1994.

[6] X. Geng, Z. Zhou, K. Smith-Miles: "Automatic age estimation based on facial aging patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.29, no.12, pp.2234-2240, 2007.

[7] FGNET: "http://www-prima.inrialpes.fr/FGnet/html/home.html."

[8] P. Phillips, H. Moon, S. Rizvi, and P. Rauss: "The feret evaluation methodology for face recognition algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.22, no.10, pp.1090-1104, 2000.

[9] A. Albert and K. Ricanek Jr: "The morph database: investigating the effects of adult craniofacial aging on automated face-recognition technology," *Forensic Science Communications*, vol.10, no.2, 2008.

[10] M. Minear and D. C. Park: "A lifespan database of adult facial stimuli," *Behavior Research Methods, Instruments, & Computers*, vol.36, no.4, pp.630-633, 2004.

[11] T. Cootes, G. Edwards, C. Taylor: "Active appearance models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.23, no.6, pp.681-685, 2001.

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

[13] J. Ylioinas, A. Hadid, X. Hong, M.Peitikainen: "Age estimation using local binary pattern kernel density estimate," *Image Analysis and Processing (ICIAP), Springer*, pp.141-150, 2013.

[14] C.-C. Ng, M.H. Yap, N. Costen, B. Li: "Will wrinkle estimate the face age?" *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp.2418-2423, 2015.

[15] E. Zhou, H. Fan, Z. Cao, Y. Jiang, Q. Yin: "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," *IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pp.386-391, 2013.

[16] C.-C. Ng, M.H. Yap, N. Costen, B. Li: "Wrinkle detection using hessian line tracking" *IEEE Access*, vol. 3, pp.1079-1088, 2015.

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