

Will Wrinkle Estimate the Face Age?

Choon-Ching Ng, Moi Hoon Yap, Nicholas Costen
School of Computing, Mathematics & Technology
Manchester Metropolitan University
United Kingdom
{choon.c.ng, m.yap, n.costen}@mmu.ac.uk

Baihua Li
Department of Computer Science
Loughborough University
United Kingdom
b.li@lboro.ac.uk

Abstract—The majority of current facial age estimation methods are based on appearance based features. However, wrinkle-based research has not been widely addressed. In this paper, we propose a novel method based on multi-scale aging patterns (MAP). These directly extract the features from local patches without extensive geometric modelling. First, we locate facial landmarks by using the Face++ detector and then normalize the face by using a linear transformation. We define a face template which consists of ten predefined wrinkle regions. Then, for each region, we detect wrinkles and construct aging patterns by using the proposed methods. Finally, the age is estimated by implementing the support vector machine for regression. The performance of the algorithms is assessed by using mean absolute error (MAE) on the benchmark database - FERET. We observe that MAP produces the lowest MAE of 4.87 on FERET compared to the benchmark algorithms. Therefore, we conclude that wrinkle could be used as a feature on face age estimation. Future work would involve improvements of the algorithm by combining other descriptors such as non-wrinkle descriptor and appearance parameters.

Keywords-Age estimation, wrinkle detection, multi-scale filter, support vector regression, FERET.

I. INTRODUCTION

Automatic facial age estimation is an important, yet largely unsolved, challenging problem. This challenge can be attributed to (i) large intra-subject variations and (ii) large inter-subject similarity. Major intra-subject variations include craniofacial growth and skin texture, meanwhile, the inter-subject similarities are gender and race. In addition, 2D facial images contain huge variations in expression, pose and illumination in the uncontrolled environment. From these variations, local-based aging features such as wrinkles, pores and spots are now beginning to receive increasing attention in the facial aging community [1], [2].

The process of age estimation attempts to label a facial 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 [3], age estimation is still a challenging problem due to three reasons [4]. First, facial age progression is uncontrolled and personalized. Such special characteristics of aging variation cannot be captured easily due to the large variations conveyed by human faces. Second,

there is no complete facial aging dataset with chronological ages. 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. For example, these are weak arguments if we say two wrinkles at the eye corner is classified as age 20, three wrinkles as 30 years old and so on. Such vague arguments will cause ambiguity if the person's age is between 20 and 30. Thus, a robust wrinkle-based representation is needed to cope with the identified problem.

The paper is organised into the following sections: Section II discusses the related work on age estimation; Section III, IV, V outline the proposed framework and methods; Section VI presents the results and discussion; finally, the conclusion is summarized in Section VII.

II. RELATED WORK

Conventionally, methods in age estimation are based on global features. Global features are computed or modelled for an entire object or scene with a large matrix or vector [5]. For facial aging research, many algorithms [4], [6]–[8] deal with appearance parameters produced by the active appearance model (AAM) [9]. AAM is a generative model which constructs the target subject's face by a set of hidden parameters [10]. Geng et al. [4] proposed an idea which is called aging pattern subspace (AGES) which makes use of AAM parameters as well. The main idea of AGES is to find the missing pattern in the aging subspace by using PCA and Expectation Maximization (EM) method. They argued that the concepts of identity and time are naturally integrated into the facial aging subspace. However, the AGES method is less representative due to the appearance model only encoding the image intensities without considering any aging characteristics such as wrinkles. Moreover, the use of PCA on appearance parameters might lose significant features because the least important variance could be either noise or aging information. As mentioned in [11], wrinkle-based features such as skin texture are more effective for face representation because they inherently contains spatial locality and orientation selectivity. These properties allow for simplicity of feature extraction and avoid the extensive modelling of AAM. Published approaches to age estimation based on wrinkle features are limited [12]–

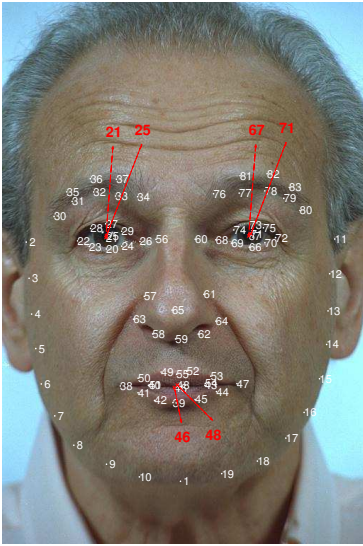


Fig. 1. Result of the Face++ detector [20]. Note that this image was redrawn from the FERET dataset [21]. White landmarks are detected by the Face++ detector and the red labels represent the center of the eyes and mouth.

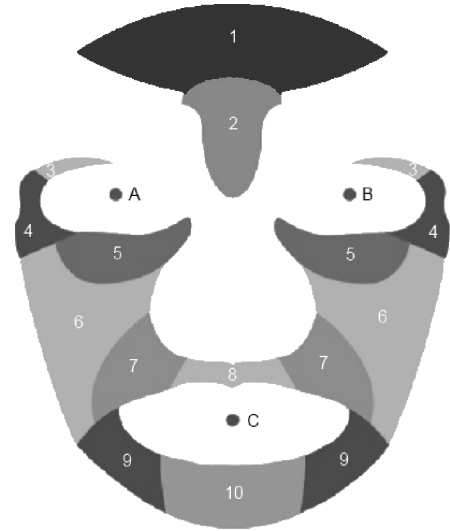


Fig. 2. Face template with ten predefined wrinkle regions and fixed coordinates of eyes and mouth. The wrinkle regions are forehead (1), glabella (2), upper eyelids (3), crow's feet (4), lower eyelids (5), cheeks (6), nasolabial grooves (7), upper lips (8), marionette (9) and lower lips (10). Symbols A, B and C are fixed coordinates used for face normalization.

[17]. Most of them were focused on age group classification instead of specific age estimation. Therefore, the capabilities of wrinkle in age estimation is not yet explored.

The current state-of-the-art method for local-based age estimation is proposed by Juha et al. [18] and Nguyen et al. [19]. The method is based on the local binary patterns (LBP) and generates feature representation through face patches. LBP is a unique pattern representation by thresholding the pixels of an image within a certain neighborhood. However, we found that the significant aging patterns such as wrinkles and spots cannot be highlighted by the LBP. Instead, the aging-like patterns will be weakened and hard to discriminate from other patterns. Our work is improving the weaknesses of LBP by highlighting the facial wrinkles which are very important for estimating age.

The skin changes associated with aging are the focus of many surgical and non-surgical procedures aimed to improve the appearance of skin [22]. Knowledge of skin histology will deepen the understanding of cutaneous changes associated with aging and will promote optimal cosmetic and functional patient outcomes. Due to these reasons, research into age estimation by using local features has gained increasing attention. Aznar-Casanova et al. [23] investigated the influence of wrinkles on facial age judgements. Their experiments were based on the types of wrinkles and quantitative contribution of wrinkles. They found that the amount of wrinkles on the perceived facial age had more influence than the types of wrinkles. Although their works sounds interesting, the whole experiment is done manually. Günay and Nabiyevev [24] proposed age estimation based on local Radon features. The idea is to transform an image pixel into an equivalent geometric Radon vector. The authors stated the proposed method achieved good results on the FG-NET dataset, but we noticed that the MAE of FERET (6.98) is worse than FG-NET (6.18). A better result should

be achieved on the FERET dataset because image resolution of FERET is higher than FG-NET. In summary, existing methods required user intervention, were not effective or had undesirable features.

In this paper, we propose a novel method for age estimation, namely multi-scale aging patterns (MAP). This work extends the wrinkle detection research by Ng et al. [2]. They proposed a hybrid Hessian filter (HHF) to detect wrinkles on forehead images. We expand the multi-scale second order local structure of a facial image into the context of age estimation by proposing a novel aging pattern. In other words, a multi-scale filter is applied to the whole face and transforms wrinkles into useful aging patterns where MAP is generated. Once the aging patterns are constructed, a sequential minimal optimization (SMO) [25] is utilized to learn and predict the age of the patterns. We tested the proposed methods on FERET dataset by using 5-folds cross validation. Both training and testing sets are disjoint.

III. AGE ESTIMATION FRAMEWORK

Facial wrinkles are created by repeated facial muscular movements and expressions [26]. Consequently, facial wrinkles have unique characteristics depending upon the frequency of movements and the effect of gravity. Many argue that facial wrinkles are highly correlated with age [23], [26]. Therefore, in this work, we explore the uniqueness of wrinkle characteristics by localizing and presenting it in a meaningful way for age estimation. A multi-scale filter is applied for assessing the geometrical structures of skin. Since wrinkles appear in different sizes, it is important to present a dynamic scale which varies within a certain range. Based on multi-level analysis of Hessian eigenvalues, the local behaviour of an image is emphasized and thus wrinkles are identified. Then, extracted

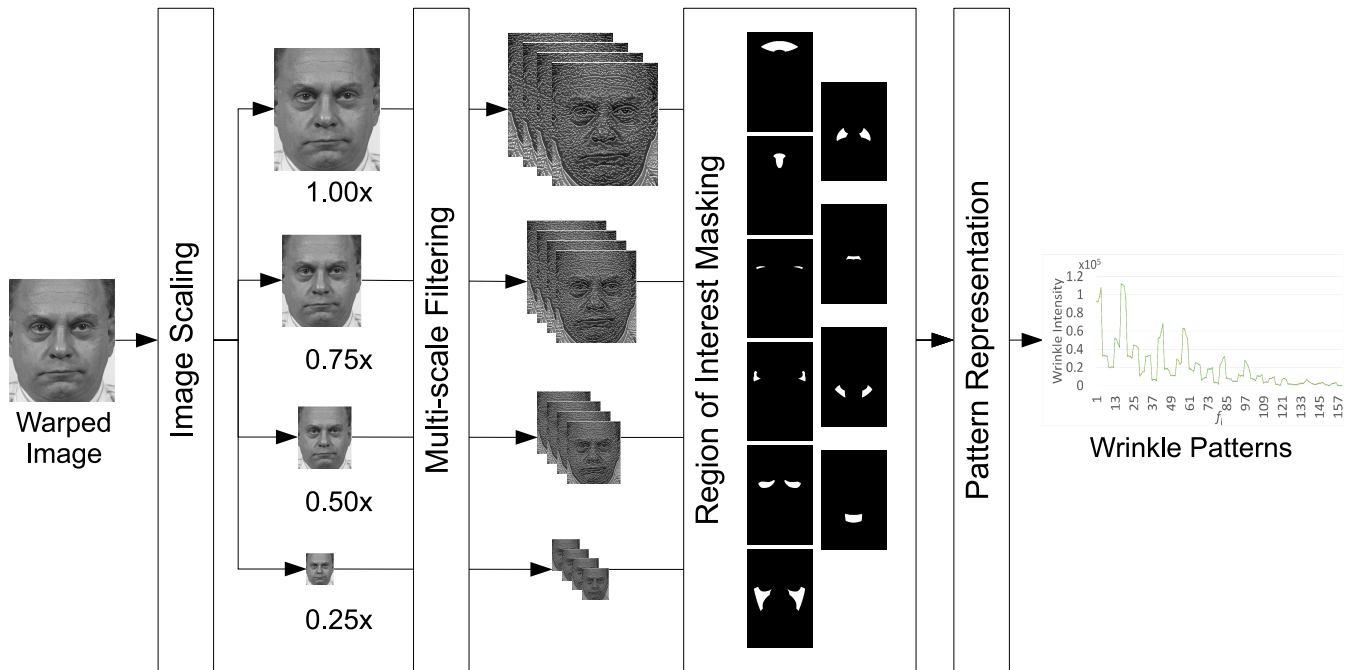


Fig. 3. Multi-scale aging patterns.

wrinkles are presented as multi-scale aging patterns and the age of a facial image is estimated through supervised learning.

A. Face Detection

First, a face image is detected by the Face++ detector [20] and a total of 83 landmarks is obtained as shown in Fig. 1. Only six landmarks are used for estimating the center of left eye, right eye and mouth, others are discarded. The center of left eye is averaged from the landmarks 21 and 25, the center of right eye is averaged from the landmarks of 67 and 71, the center of mouth is averaged from the landmarks 46 and 48. Symbols A, B and C refer to the center of left eye, center of right eye and the center of mouth, respectively, as shown in Fig. 2. Due to the limitations of the Face++ detector, we expect a small number of failed detections which will be identified manually.

B. Face Normalization

Fig. 2 illustrates the face template or mask with ten predefined wrinkle regions and fixed coordinates for eyes and mouth. It was cropped and scaled for illustration purposes and the original size of this image is 512 x 768 pixels which is the same size as FERET images. These regions have been identified and created manually through the review [23], [26]. The margins of top, bottom, left and right are 70.5, 209.5, 50.5 and 56.5 pixels, respectively. The distance between A and B is 218 pixels, A and C is 236.6 pixels, B and C is 237.49 pixels. Numbers 1 to 10 mean the wrinkle regions of forehead, glabella, upper eyelids, crow's feet (or eye corners), lower eyelids (or eyebag), cheeks, nasolabial grooves (or nasolabial folds), upper lips, marionette and lower lips; the corresponding areas are 15745, 6437, 1475, 4169, 9417, 21759, 9216, 2298,

7376 and 7403 pixels; the corresponding notations are \mathcal{Z}_1 to \mathcal{Z}_γ where \mathcal{Z}_i is a binary image and $\gamma = 10$. Due to the area sizes being fixed, a standard feature vector will be produced either in training or testing data. The center coordinates of left eye, right eye and mouth are (145, 247), (363, 247) and (254, 457), respectively. All regions are used for constructing the aging patterns. Based on the three points (symbols A, B and C), each face image is normalized to the mask (as shown in Fig. 2) by using pieewise affine warping [9].

Procrustes analysis [27] is applied in between the input face and mask in order to determine the linear transformation. Given a set of landmarks $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, we compute the transformation [9] based on Procrustes analysis [27] so that sum of distance \hat{d} of face shape L to the mean face \bar{L} is minimised as $\hat{d} = \sum |L - \bar{L}|^2$. Procrustes Analysis is a method of comparing two sets of shapes by fitting one data set to another by rotation, translation, uniform scaling, and reflection. The transformed warped image of L is then given by $I = \hat{f}(L)$ where \hat{f} is an affine geometric transformation function [9]. For each face in the training set, we compute the aging patterns $\mathcal{F} = \mathcal{P}(I)$, where $\mathcal{P}(\cdot)$ denotes the proposed method MAP as described in Section IV. Finally, the age is estimated by using the regression function $\text{age} = \text{SMO}(\mathcal{F})$ as described in Section V.

In the next step, wrinkles patterns are constructed by using the proposed method MAP. Once the normalized face image is filtered by the multi-scale filter [2], extracted outputs are wrinkles. Then, ROI masks are applied on the face and the intensity summation is applied on each region in order to produce the aging patterns. Section IV shows these methods in detail. Finally, the age of each aging pattern is predicted by

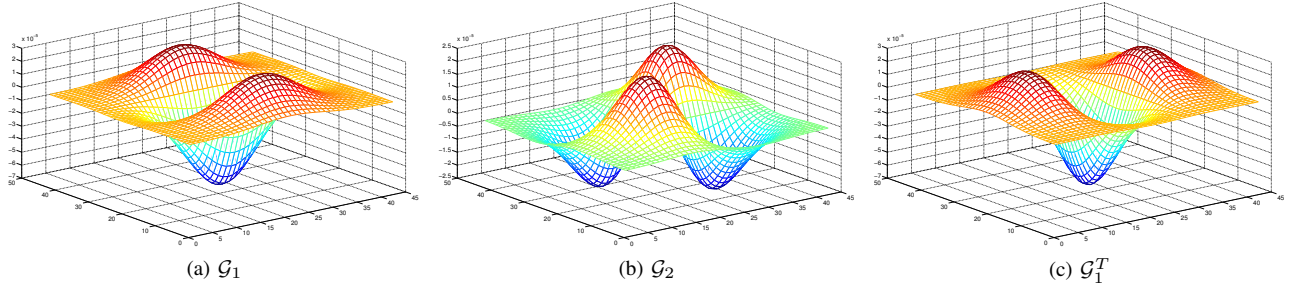


Fig. 4. Gaussian kernels \mathcal{G}_1 , \mathcal{G}_2 and \mathcal{G}_1^T where $\sigma = 7$.

the SMO. Section V describes the SMO in detail.

IV. MULTI-SCALE AGING PATTERNS

Fig. 3 demonstrates the proposed multi-scale aging patterns. It consists of four steps: image scaling, multi-scale filtering, ROI masking and aging pattern representation.

A. Image Scaling

Given a warped image \hat{I} , it is scaled with different ratios s and denoted as $I = \hat{I}(s)$ where $s \in \{s_1, s_2, \dots, s_\alpha\}$ and α is the total number of scales. We set $s \in \{1.00, 0.75, 0.50, 0.25\}$. Note that both the warped image and ROI mask image are the same size, therefore they will be scaled accordingly.

B. Multi-scale Filtering

As reported in [2], the directional gradient $(\partial I/\partial x, \partial I/\partial y)$ is computed from the greyscale image I . Let $(\partial I/\partial y)$ be denoted as \mathcal{I} , therefore \mathcal{I} emphasizes a horizontal variation and is used as the input for calculating the Hessian matrix \mathcal{H} . The Hessian matrix \mathcal{H} is the output of the second derivative at $\sigma \in \{1, 3, 5, 7\}$ [2]. Each approximates the convolution of \mathcal{I} by Gaussian kernels \mathcal{G}_1 and \mathcal{G}_2 where it measures the contrast at the selective scale in the direction of the derivative. Fig. 4 shows the Gaussian kernels \mathcal{G}_1 , \mathcal{G}_2 and \mathcal{G}_1^T where $\sigma = 7$. Since wrinkles are similar to the patterns of ridge and valley, the Gaussian kernels are designed in the same way. In order to determine the local likelihood that a wrinkle is present, eigenvalues of the Hessian matrix is assessed. It highlights the texture orientation and discard noisy patterns [28]. Finally, curvilinear likeliness \mathcal{E} of eigenvalues is calculated. It is used for generating the initial wrinkle mask as

$$\mathcal{B}_{s,\sigma}(x,y) = \begin{cases} 0 & \text{if } \mathcal{E}_{s,\sigma}(x,y) > 0 \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

C. ROI Masking

Assume the wrinkle image J is masked as

$$\mathcal{J}(x,y) = \begin{cases} I(x,y) & \text{if } \mathcal{B}(x,y) \cap \mathcal{Z}(x,y) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where \mathcal{Z} is the specific region of template mask as discussed in section III.

D. Pattern Representation

Let $\text{MAP} = \{f_1, f_2, \dots, f_\psi\}$, where $\psi = \alpha \times \gamma \times \varphi$. Note that, in this work, we set total image scale $\alpha = 4$, total wrinkle regions $\gamma = 10$ and total filter scales $\varphi = 4$. In total, 160 features are produced for each image. The feature f_i is the wrinkle intensity and it is defined as the summation of \mathcal{J}_i which given by

$$f_i = \sum_{x=1}^{wt} \sum_{y=1}^{ht} \mathcal{J}_i(x,y) \quad (3)$$

where wt and ht are the width and height of \mathcal{J} .

V. SUPPORT VECTOR MACHINE FOR REGRESSION

Support vector machines (SVM) were originally developed for the classification problem by Boser et al. [29]. This technique was built on the principle of structural risk minimization (SRM). Now, with the introduction of the insensitive loss function, SVM has been extended to solve nonlinear regression estimation [25]. By using kernel functions, SVM plays a role in mapping the data to a high dimensional feature space and then finds a linear separating hyperplane with the maximal margin in this high dimensional space.

In this paper we used an improved sequential minimal optimization for SMO [25] with the normalized polynomial kernel to optimise prediction. As pointed by Graf et al. [30], the normalization of vectors in input space is not as good as in feature space due to the prediction data not being normalized. Therefore, normalization in the feature space is a better solution for both training and prediction. Unlike in linear programming methods [31] where normalization acts on rows and columns of a design matrix only, normalization of kernel functions can be considered as a simultaneous rescaling of rows and columns to obtain a matrix with all diagonal entries set to one [30].

Let $\vec{x}, \vec{y} \in \mathfrak{R}^N$ denote input vectors of SMO, the normalized polynomial kernel κ is defined by (4) and the polynomial kernel is denoted as (5) where this kernel consists of all monomials up to degree d . The kernel with $d = 1$ is the linear kernel. The classifier's flexibility increases when d is getting higher. However, if this parameter is too large, overfitting will occur [32].

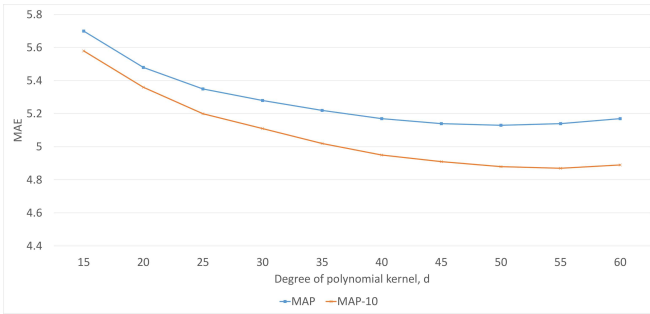


Fig. 5. MAE of different exponent d on the FERET database. MAP-10 means MAP with 10-cross validation.

TABLE I

MAE OF DIFFERENT AGE GROUPS. MAP-10 MEANS MAP WITH 10-CROSS VALIDATION.

Algorithm	10-19	20-29	30-39	40-49	50-59	60-70
Total images	38	950	629	474	209	66
FACE++ [20]	11.08	7.13	5.92	8.05	6.94	3.47
MAP	12.11	4.89	3.54	4.87	8.33	11.39
MAP-10	9.93	3.94	4.02	4.37	5.48	8.91

$$\kappa(\vec{x}, \vec{y}) = \frac{\varphi(\vec{x}, \vec{y})}{\sqrt{\kappa(\vec{x}, \vec{x}) \kappa(\vec{y}, \vec{y})}} \quad (4)$$

$$\varphi(\vec{x}, \vec{y}) = (\langle \vec{x}, \vec{y} \rangle + 1)^d \quad (5)$$

In this work, we implemented age estimation experiments by using the WEKA toolbox [33]. The regression algorithm is SMO and the default parameters are as follows: $c = 1.0$, ‘filterType’ is ‘Normalize training data’, kernel is ‘Normalized-PolyKernel’, ‘regOptimizer’ is ‘RegSMOImproved’, epsilon for round-off error is $1.0e^{-12}$, the epsilon parameter of the epsilon insensitive loss function is 0.001, seed is 1, tolerance parameter used for checking stopping criterion is 0.001 and ‘useVariant1’ is true.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed methods is tested on the FERET dataset [21]. FERET is a comprehensive database with 2366 images of 994 subjects that presents multiple problems related to face recognition such as illumination variations, pose variations, facial expressions, etc. In addition, it consists of a few hundred age-separated face images of subjects with the age difference of 18 months or more and the age range is between 10 and 70. Two metrics have been proposed in the literature for quantifying the performance of age estimation methods, MAE and cumulative score (CS) [34], [35]. The MAE is defined as the average of the absolute errors between estimated age and the ground truth, $MAE = \sum_{k=1}^N (|\hat{l}_k - l_k| / N)$, where l_k is the ground truth age for the k -th test image, \hat{l}_k is the estimated age, and N is the total number of test images. The CS [36], is defined as $CS(j) = N_{error \leq j} / N \times 100\%$, where $N_{error \leq j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years. $CS(j)$ can be viewed as the classification accuracy, which, like in face recognition,

TABLE II

COMPARISON OF MAE RESULTS ON THE FERET DATABASE. NOTE THAT FACE++ RESULTS WERE CALCULATED BY REPLACING THE FAIL DETECTION WITH GROUND TRUTH. MAP-10 MEANS MAP WITH 10-CROSS VALIDATION.

Algorithm	MAE
Radon [24]	6.98
Face++ [20]	6.94
Proposed Method, MAP	5.13
Proposed Method, MAP-10	4.87

might be more important than the average performance MAE for practical applications. Here, we use both CS and MAE as metrics since different methods, databases, and systems may be too biased or unbalanced for evaluation.

Fig. 5 shows MAE results of different degree d of polynomial kernel on the FERET dataset. We noticed that MAP and MAP-10 hit the lowest MAE of 5.13 when $d = 50$ and MAE of 4.87 when $d = 55$. Table I presents the MAE of different age groups. FERET consists of six age group - 10-19, 20-29, 30-39, 40-49, 50-59, 60-70 and the total images of each group are 38, 950, 629, 474, 209 and 66, respectively. From the experiment, MAP achieved the lowest MAE of 3.54 on age group of 30-39 while MAP-10 hit the lowest MAE of 3.94 on age group of 20-29. Both groups contain the largest amount of images in FERET. Although MAP achieved the lowest MAE compared to MAP-10, but MAP-10 performed better in overall result. It could be due to 10-folds cross validation used more training images than 5-folds cross validation. Age group of 10-19 remains the high MAE in both experiments because it is very hard to extract wrinkle-based features for children and teenagers. FACE++ achieved the lowest MAE on age group of 60-70. It would be interesting to explore the potential features of FACE++ on older group. Table II presents the comparison of MAE results on the state-of-the-art methods. MAP is a superior method in extracting the local-based aging features. Therefore, it has the lowest MAE, at 4.87, and this outperformed the Radon method [24] and Face++ detector [20], at 6.98 and 6.94. Based on these results, we presented that age estimation using wrinkle-based features is workable on high resolution images.

The idea of this work is based on the intensity distribution of facial wrinkles on the predefined wrinkle regions. It is useful to identify a simple yet powerful pattern from the same age group. However, we noted four issues that affect the performance of age estimation. First is the mask of wrinkle regions. Although this has been set to the size of the mask, we noticed that some of the faces cannot fit the mask very well due to large variations in the cranial ratio and occlusions such as forehead hair and moustaches. Second, there are false detections of landmarks in a few images detected by the Face++ detector. Although they were identified and corrected manually, the estimation result was affected. Third, cosmetic treatment and facial expression affect the wrinkle appearance on face. It would be interesting to see such impact on face age estimation in the future. Finally, a robust wrinkle-based features is highly depends on the wrinkle extraction method. Although this method is limited to high resolution images, additional fusion concerning features of

hierarchical age estimation can be added in the near future.

VII. CONCLUSION

This paper has proposed a novel method for facial age estimation based on local-based wrinkle features. By deriving wrinkles with the multi-scale filters across 10 face regions, wrinkle-based aging patterns are generated. They are used to train the aging classifier and estimate the age of faces. A novel algorithm named MAP is proposed in this paper to represent wrinkles into a representative pattern. They are tested on FERET aging face databases. Experimental results show the evidence that wrinkles is able to predict face age on high resolution images. Further investigation of a hybrid method which combines global and local features into an aging pattern can be used to enhance the proposed methods. This method can compensate for defects found in individual use of global or local features.

ACKNOWLEDGMENT

This work was supported by a PhD Studentship from Manchester Metropolitan University. The authors would like to thanks Phillips et al. [21] for the FERET dataset and Zhou et al. [20] for the FACE++ detector.

REFERENCES

- [1] N. Batool and R. Chellappa, "Detection and inpainting of facial wrinkles using texture orientation fields and Markov random field modeling," *IEEE Trans. On Image Processing*, vol. 23, no. 9, pp. 3773–3788, 2014.
- [2] C.-C. Ng, M. Yap, N. Costen, and B. Li, "Automatic wrinkle detection using hybrid hessian filter," in *Computer Vision – ACCV 2014*, ser. Lecture Notes in Computer Science, D. Cremers, I. Reid, H. Saito, and M.-H. Yang, Eds. Springer International Publishing, 2015, vol. 9005, pp. 609–622.
- [3] Y. Kwon and N. da Vitoria Lobo, "Age classification from facial images," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1994, pp. 762–767.
- [4] X. Geng, Z. Zhou, and 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.
- [5] D. A. Lisin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, "Combining local and global image features for object class recognition," in *IEEE Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2005, pp. 47–47.
- [6] C. Chen, Y. Chang, K. Ricanek, and Y. Wang, "Face age estimation using model selection," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2010, pp. 93–99.
- [7] 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 (CVPR)*. IEEE, 2011, pp. 585–592.
- [8] W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognition*, vol. 46, no. 3, pp. 628 – 641, 2013.
- [9] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [10] X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 2, pp. 145–158, 2010.
- [11] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 1028–1037, 2011.
- [12] Y. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.
- [13] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of age groups based on facial features," *Tamkang Journal of Science and Engineering*, vol. 4, no. 3, pp. 183–192, 2001.
- [14] R. Iga, K. Izumi, H. Hayashi, G. Fukano, and T. Ohtani, "A gender and age estimation system from face images," in *SICE 2003 Annual Conference*, vol. 1. IEEE, 2003, pp. 756–761.
- [15] H. Takimoto, Y. Mitsukura, M. Fukumi, and N. Akamatsu, "A design of gender and age estimation system based on facial knowledge," in *International Joint Conference SICE-ICASE*. IEEE, 2006, pp. 3883–3886.
- [16] H. Fukai, H. Takimoto, Y. Mitsukura, and M. Fukumi, "Apparent age estimation system based on age perception," in *SICE*, 2007, pp. 2808–2812.
- [17] C. Shan, "Learning local features for age estimation on real-life faces," in *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*. ACM, 2010, pp. 23–28.
- [18] J. Ylioinas, A. Hadid, X. Hong, and M. Pietikäinen, "Age estimation using local binary pattern kernel density estimate," in *Image Analysis and Processing-ICIAP 2013*. Springer, 2013, pp. 141–150.
- [19] D. T. Nguyen, S. R. Cho, and K. R. Park, "Human age estimation based on multi-level local binary pattern and regression method," in *Future Information Technology*. Springer, 2014, pp. 433–438.
- [20] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," in *IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*. IEEE, 2013, pp. 386–391.
- [21] 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.
- [22] J. Khavkin and D. Ellis, "Aging skin: histology, physiology, and pathology," *Facial plastic surgery clinics of North America*, vol. 19, no. 2, p. 229, 2011.
- [23] J. Aznar-Casanova, N. Torro-Alves, and S. Fukusima, "How much older do you get when a wrinkle appears on your face? modifying age estimates by number of wrinkles," *Aging, Neuropsychology, and Cognition*, vol. 17, no. 4, pp. 406–421, 2010.
- [24] A. Günay and V. V. Nabyev, "Age estimation based on local radon features of facial images," in *Computer and Information Sciences III*. Springer, 2013, pp. 183–190.
- [25] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Trans. on Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000.
- [26] 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, no. 6, pp. 1262–1281, 2011.
- [27] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339, 1991.
- [28] A. F. Frangi, "Three-dimensional model-based analysis of vascular and cardiac images," PhD Thesis, University Medical Center Utrecht, 2001.
- [29] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM, 1992, pp. 144–152.
- [30] A. B. Graf, A. J. Smola, and S. Borer, "Classification in a normalized feature space using support vector machines," *IEEE Trans. on Neural Networks*, vol. 14, no. 3, pp. 597–605, 2003.
- [31] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optimization Methods and Software*, vol. 1, no. 1, pp. 23–34, 1992.
- [32] A. Ben-Hur and J. Weston, "A users guide to support vector machines," in *Data Mining Techniques for the Life Sciences*. Springer, 2010, pp. 223–239.
- [33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [34] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.
- [35] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.
- [36] G. Guo, Y. Fu, C. Dyer, and T. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.