EXTRACTING AGE INFORMATION FROM LOCAL SPATIALLY FLEXIBLE PATCHES

Shuicheng Yan[1], *Ming Liu*[2], *and Thomas S. Huang*[2]

[1]Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[1]Beckman Institute for Advanced Science and Technology, UIUC, USA
eleyans@nus.edu.sg

ABSTRACT

Motivated by the fact that age information can often be observed from local evidence on the human face, we contribute to the age estimation problem in two aspects. On the one hand, we present a new feature descriptor, called spatially flexible patch (SFP), which encodes the local appearance and position information simultaneously. SFP has the potential to alleviate the problem of insufficient samples owing to that SFPs similar in appearance yet slightly different in position can still provide similar confidence for age estimation. One the other hand, the SFP associated with age label is modeled with Gaussian Mixture Model, and then age estimation is conducted by maximizing the sum of likelihoods from all the SFPs associated with the hypothetic age. Experiments are conducted on the YAMAHA database with 8,000 face images and ages ranging from 0 to 93. Compared with the latest reported results, our new algorithm brings encouraging reduction in mean absolute error for age estimation.

Index Terms— Age Estimation, Gaussian Mixture Model, Spatially Flexible Patches.

1. INTRODUCTION

Age is a significant characteristic of humans, yet only a few works have been dedicated to the age estimation problem [6] [7] [10]. This problem is challenging due to the complex effects caused by living conditions, cosmetics usage, personal specialties, gender differences, and so on.

1.1. Backgrounds

Kown et al. [7] presented a method for age classification based on cranio-facial development theory and skin wrinkle analysis, with human faces finally classified into three groups: babies, young adults and senior adults. Hayashi et al. [6] proposed to classify age and gender based on the wrinkle and geometry relationships between different parts of a face, with the human ages divided into multiple groups at the five year intervals. Lanitis et al. [8] utilized Active Appearance Models [2] to extract the combined shape and texture parameters, and then compared the age estimation accuracies of the algorithms including simple quadratic fitting, shortest distance classifier, and Neural Network. Geng et al. [5] [4] proposed to conduct age estimation by modeling the statistical properties of aging patterns, namely a sequence of personal facial age images, based on the assumption that multiple images of different ages are available for each person. Recently, Yan et al. [12] proposed an algorithm for age estimation with the consideration of uncertainty of the age labels.

1.2. Motivations and Solutions

Research on age estimation is still at a preliminary stage, but has attracted much attention in the past few years. A detailed comparison of the previous algorithms is summarized in Table 1. In this work, we tackle the age estimation problem on the challenging YAMAHA database with 8,000 images, 1,600 subjects, and ages ranging from 0 to 93.

As listed in Table 1, most previous research was based on holistic appearance related features. However, an intuitive observation is that the age information can be even obtained only from a local patch on the face, and so it is natural to encode information by patches instead of holistic appearance. Another motivation for patch based age estimation is as follows. Due to the varieties of races, the geometric structures are different for different subjects, and there often does not exist exact semantic pixel-wise correspondence between faces images. Patch based representation can solve this geometric inhomogeneity issue.

In this paper, we contribute to the age estimation problem from two aspects. On the one hand, we present a new feature descriptor for face representation. A face image is characterized as an ensemble of spatially flexible patches, which encode the local appearance and relative position information simultaneously. The advantage of spatially flexible patch (SFP) is that SFPs with similar appearance yet slightly different positions can still provide similar confidence for age estimation. Moreover, compared with direct patch without position information, SFP well utilizes the spatial information in a soft manner and without the requirement that pixels should be aligned with fixed correspondence. On the other hand, we utilize the Gaussian Mixture Model (GMM) [3] for charac-

Table 1. A summary of the information on database size, age precision, and main features used for different age estimation algorithms. Note that *GR* here means the features describing the geometric relationship between different facial parts and *Wk* means the wrinkle features.

Algorithm	*Faces*	*Subjects*	*Precision*	*Features*
Kown [7]	47	-	Three groups	GR+Wk
Hayashi [6]	300	-	5y/group	GR+Wk
Lanitis [8]	400	40	0-35y	Appearance
Geng [5]	1002	82	0-69y	Appearance
Yan [12]	8,000	1,600	0-93y	Raw Image

terizing the distribution of the aged SFP, namely the variable combining the features from SFP and the associated age label. Finally, the age estimation is conducted by comparing the sum of the likelihoods from all the aged SFPs associated with the hypothetic age.

2. SPATIALLY FLEXIBLE PATCHES

For the human age estimation problem, the image set for model training is denoted here as a matrix $X = [x_1, x_2, \ldots, x_N], x_i \in \mathbb{R}^m$, where N is the image number and m is the feature dimension. The age label for the image x_i is denoted as l_i and $l_i \geq 0$. The task is to predict the age label of a new image x.

As mentioned above, most previous algorithms for age estimation are based on holistic appearance features. This representation greatly limits the robustness of the algorithms to illumination, pose, and expression variation. In this work, we propose a descriptor called spatially flexible patch (SFP) as demonstrated in Figure 1 for encoding the information of a face in a local manner. SFP integrates both local appearance and position information. For a position within the image plane $p = (p_x, p_y)^T$, its corresponding SFP for sample x_i is represented as

$$P(x_i, p) = \begin{bmatrix} x_i(R(p)) \\ p \end{bmatrix}, \tag{1}$$

where $R(p)$ means the index set of the pixels within the rectangle centered by position p.

For the intensity feature of each patch, we first remove the mean of the intensity values and then normalize the appearance with unit variance, and finally we use the DCT transform for extracting the local features. In this work, the position p is densely sampled for all possible positions.

The advantages of SFP descriptor are as follows: 1) It is flexible. It is often the case in real applications that the face image cannot be pixel-wisely matched with explicit semantics due to different geometric structures of different subjects, and also there may exist misalignments when cropping faces. In these cases, SFP can well alleviate the issue, and two SFPs with similar appearance yet with slightly different positions can still be matched and provide similar confidence for age estimation. 2) It is robust. Each SFP is only portion of the holistic appearance, and hence SFPs are relatively robust to occlusion as an ensemble. Also, the final SFP appearance feature is normalized, so it is robust to illumination variation. 3) It integrates both appearance and position information. Conventional patch based algorithms [9] ignore the position information which may also be useful for classification since human face is a kind of well structured object.

3. GMM MODELING AND INFERRING BASED ON AGED SPATIAL FLEXIBLE PATCHES

After feature extraction, each face image is represented as an ensemble of SFPs, and we use these SFPs as features for age estimation. Each SFP will be considered as a unit for age estimation, and the SFP is combined with the age label to constitute the variable for modeling. Here, we assume that

$$z = \begin{bmatrix} P(x_i, p) \\ l_i \end{bmatrix}, \tag{2}$$

and z represents the aged SFP.

In this work, we use the Gaussian Mixture Model for characterizing the distribution of the variable z, namely,

$$p(z|\lambda) = \sum_{k=1}^{K} \omega_k p_k(z). \tag{3}$$

The density is a weighted linear combination of K unimodal Gaussian densities, $p_k(z)$, each parameterized by a mean vector, μ_k, and a covariance matrix, Σ_k, namely,

$$p_k(z) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_k|^{\frac{1}{2}}} exp\{-\frac{1}{2}||z - \mu_k||^2_{\Sigma_k^{-1}}\}. \tag{4}$$

As conventionally, we use the Expectation Maximization approach for pursuing of the parameters. In this paper, we constrain the covariance matrices to be diagonal for computational efficiency.

When a new image comes, we first extract all the SFPs as

$$P = \{P(x, p), \forall p\}, \tag{5}$$

and its age information can be predicted by maximizing the sum of likelihoods from all the SFPs,

$$\arg\max_l \sum_p p(\begin{bmatrix} P(x, p) \\ l \end{bmatrix} |\lambda). \tag{6}$$

The objective function is nonlinear and theoretically there does not exist closed-form solution. For the age estimation problem, the estimation accuracy is enough when up to integer level, and hence we simply the problem as

$$\arg \max_{l \in \{0,1,2,\cdots,M\}} \sum_p p(\begin{bmatrix} P(x, p) \\ l \end{bmatrix} |\lambda), \tag{7}$$

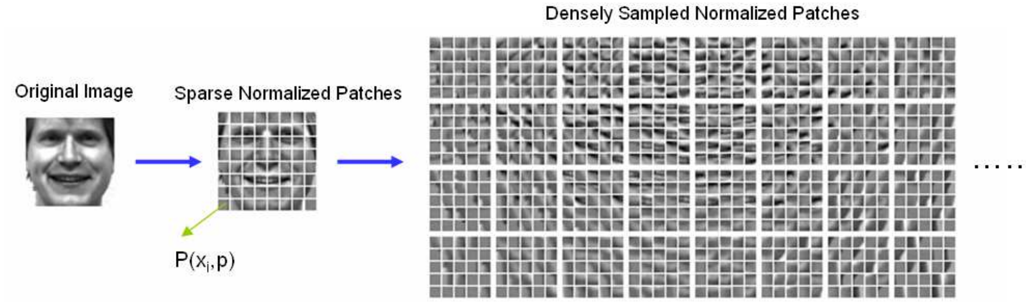


Fig. 1. Process of extracting spatially flexible patches.

where M is the largest possible age, and in this work M is set as 93 since the largest age of the database is 93. As shown in Figure 2, the distribution of $\sum_p p(\left[\begin{array}{c} P(x,p) \\ l \end{array}\right] |\lambda)$ over l is nearly unimodal around the range with the largest likelihood, so we can also use multi-scale search approach for pursuing more exact age information if required.

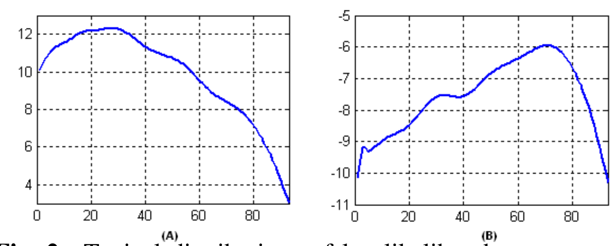


Fig. 2. Typical distributions of log-likelihood outputs over different ages for certain new face image: (A) a new image from a young person, and (B) a new image from a senior person.

4. EXPERIMENTS

In this section, we compare our proposed spatially flexible patch (SFP) based algorithm with the state-of-the-art algorithms for age estimation. The YAMAHA aging database is used for the experiments. FG-NET [1] aging database is also very popular, but it is commonly evaluated based on the shape parameter and texture parameter from the active appearance model [2], and hence we did not test our algorithm on it.

4.1. Experimental Configurations

As in [12], our algorithm was also compared with the traditional regression algorithms Quadratic Models (QM) and supervised Neural Networks [8]; and all the parameters are set as in [12]. Also we compared the proposed the algorithm called RUN in [12], which is the latest reported algorithm for age estimation.

Two measures are used to evaluate the algorithmic performance. The first one is the Mean Absolute Error (MAE) criterion used in [8] [5]. MAE is defined as the average of the absolute errors between the estimated labels and ground truth labels, *i.e.*, $MAE = \sum_{i=1}^{N_t} |\hat{a}_i - a_i|/N_t$, where $\hat{a}_i$ is the estimated age for the i-th testing sample, a_i is the ground truth age and N_t is the number of testing images. Another popular measure is the cumulative score [5]: $CumScore(\theta) = N_{e\leq\theta}/N_t \times 100\%$, where $N_{e\leq\theta}$ is the number of samples on which the absolute errors are not higher than θ.

Yamaha[1] aging database contains 8000 Japanese facial images of 1600 persons with ages ranging from 0 to 93. Each person has 5 images and the YAMAHA database is divided into two parts with 4000 images from 800 males and another 4000 images for 800 females. Our experiments are carried out separately on female and male subsets. For each subset, 1000 images are randomly selected for model training while the remaining 3000 samples are used for testing, and configurations are the same as in [12].

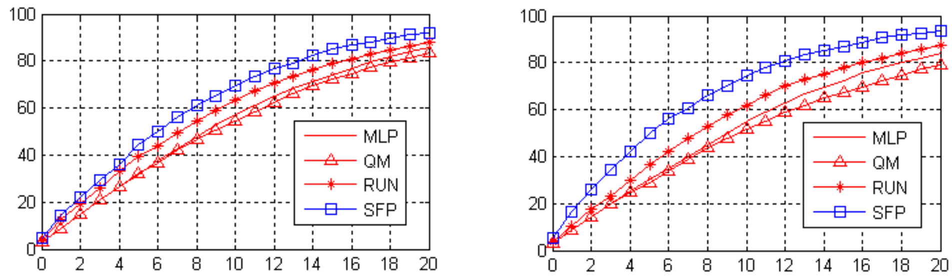
[1]To protect the portrait rights of the participants, sample images of the YAMAHA face database are not shown here.

4.2. Age Estimation Results

Figure 3 displays the cumulative scores of different algorithms, and Table 2 lists the detailed MAEs of different algorithms. From these results, we can have the observation: our SFP based algorithm reaches the lowest MAEs across both experiments, and averagely about 1.9 years' (or 19%) reduction in mean absolute error is achieved.

5. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, the age estimation problem was tackled with a framework by combining spatially flexible patches and GMM modeling of of the aged SFPs. Advantages of this framework include the ability to handle the issues of geometric misalignment and noisy patches. Encouraging experimental results have been achieved on the YAMAHA aging database by comparing with the latest reported results in literature. One future direction of this work is to conduct patch selection before the GMM modeling, and another possible direction is to explore different features for better characterizing the age information



(a) Female@Yamaha. (b) Male@Yamaha.

Fig. 3. Cumulative scores of the age estimation for QM, MLP, RUN [12] and SFP based algorithm at error levels from 0 to 20 years for different configurations.

Table 2. MAEs of different algorithms on YAMAHA and over different age ranges.

Female@Yamaha					Male@Yamaha				
Range	SFP	RUN [12]	QM	MLP	Range	SFP	RUN [12]	QM	MLP
0-9	7.20	11.21	11.97	14.33	0-9	4.51	9.86	13.42	14.08
10-19	4.55	6.23	9.58	8.85	10-19	5.43	7.52	10.33	9.46
20-29	4.79	7.95	9.29	9.70	20-29	6.53	8.85	10.21	9.35
30-39	8.43	8.17	9.85	9.66	30-39	8.27	7.76	9.35	8.60
40-49	9.49	8.64	10.45	8.78	40-49	11.62	8.67	11.71	9.10
50-59	7.42	9.43	10.15	9.53	50-59	10.44	11.10	13.38	10.08
60-69	12.12	11.12	13.49	10.88	60-69	7.50	12.49	15.99	13.44
70-93	14.27	15.56	19.66	16.52	70-93	8.21	16.60	20.44	19.69
Average	**8.53**	9.79	11.80	11.03	Average	**7.82**	10.36	13.10	11.72

within each SFP.

6. REFERENCES

[1] The fg-net aging database:http://sting.cycollege.ac.cy/∽alan-itis/fgnetaging/index.htm.

[2] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no, 6, pp. 681–685, 2001.

[3] K. Fukunaga. Introduction to statistical pattern recognition. *Academic Press, second edition*, 1991.

[4] X. Geng, Z. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press.

[5] X. Geng, Z. Zhou, Y. Zhng, G. Li, and H. Dai. Learning from facial aging patterns for automatic age estimation. *Proceedings of ACM Multimedia'06*, pp.307–316, 2006.

[6] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu. A method for estimating and modeling age and gender using facial image processing. *Seventh International Conference on Virtual Systems and Multimedia*, pp. 439–448, 2001.

[7] Y. Kwon and N. Lobo. Age classification from facial images. *Computer Vision and Image Understanding*, 74(1):1–21, 1999.

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

[9] S. Lucey and T. Chen, "Learning Patch Dependencies for Improved Pose Mismatched Face Verification," In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 909-915, 2006.

[10] V. Nallure, J. Ye, and S. Panchanathan. Biased Manifold Embedding: A framework For Peson-Independent Head Pose Estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[11] D. Rumelhart, G. Hinton and R. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[12] S. Yan, H. Wang, T. Huang, and X. Tang. Auto-Structured Regressor from Uncertain Labels. *International Conference on Computer Vision*, 2007.