Face Age Estimation
Using Patch-based Hidden Markov Model Supervectors

Xiaodan Zhuang, Xi Zhou, Mark Hasegawa-Johnson, and Thomas Huang
Beckman Institute, Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, U.S.A.
{xzhuang2,xizhou2,jhasegaw,t-huang1}@uiuc.edu

Abstract

Recent studies in patch-based Gaussian Mixture Model (GMM) approaches for face age estimation present promising results. We propose using a hidden Markov model (HMM) supervector to represent face image patches, to improve from the previous GMM supervector approach by capturing the spatial structure of human faces and loosening the assumption of identical face patch distribution within a face image. The Euclidean distance of HMM supervectors constructed from two face images measures the similarity of the human faces, derived from the approximated Kullback-Leibler divergence between the joint distributions of patches with implicit unsupervised alignment of different regions in two human faces. The proposed HMM supervector approach compares favorably with the GMM supervector approach in face age estimation on a large face dataset.

1. Introduction

Human face age estimation can provide useful information to various applications including electronic consumer relation management and demographic data collection. This is a challenging problem due to the complex variation including cosmetics usage, personal specialties, living conditions, gender and ethnic differences. The research community has expressed increasing interest and developed various approaches for age estimation [1, 2, 3, 4, 5].

While most previous algorithms for age estimation are based on holistic image features, some recent studies using patch-based Gaussian Mixture Models (GMM) present exciting results, with the merits of robustness to image occlusion and misalignment. In [7] GMMs are used to estimate the posteriors of different ages. Yan et al [12] propose multiple modules to boost the age estimation performance, in which a patch-based GMM supervector approach alone outperforms other previous works on two different face age datasets.

Although these studies present promising results, they have some intrinsic problems. First, an image is represented as an ensemble of orderless patches, thus the spatial structure is discarded in the models. Second, the distribution of the patches from an image is modeled by one single probability density function with the assumption that all these patches are identically distributed. Third, the distance between two faces is calculated without distinguishing different parts of a face.

To tackle the above problems of patch-based GMM approaches for age estimation, we propose to use a Hidden Markov Model (HMM) for unsupervised segmentation of the human face into a few regions and model the patch distribution of each region seperately. Patch-based HMM supervectors are derived such that their pairwise Euclidean distances approximate Kullback-Leibler divergence between corresponding patch distributions of the two images. This approach loosely captures the spatial structure of human faces, loosens the assumption of identical face patch distribution by modeling the regional patches seperately, and implys unsupervised alignment of different regions in different human faces.

Age estimation experiments are carried out on a face age dataset, comparing the proposed patch-based HMM supervector approach with the GMM supervector approach. In our experiments, the proposed approach outperforms the GMM supervector approach[12], which was shown to perform the best in the literature.

2 Framework

The diagram of our face age estimation system is presented in Figure 1. Each image is first represented by densely sampled and partially overlapping patches. Then we use an HMM for unsupervised segmentation of the human face, each region corresponding to one hidden state. The joint distribution of patches in each image is approximated by the regional patch distributions, i.e., the state-dependent observation distributions. For each image, the proposed patch-based HMM supervector is calculated. The pairwise Euclidean distances between these supervectors characterize the difference between the images, and are used in a nearest centroid classifier to classify a test image into an age class, each corresponding to an integer age.

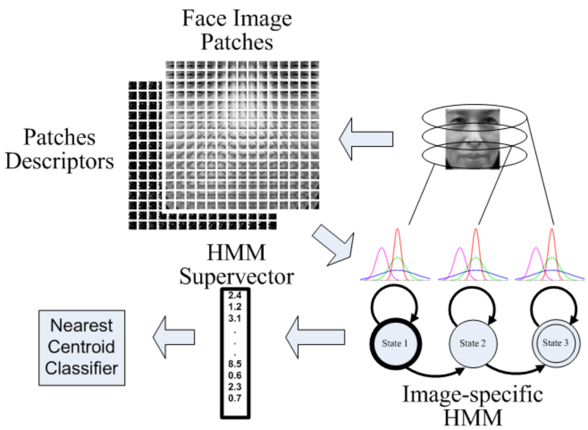


Figure 1. Face age estimation using patch-based HMM supervectors.

3 HMM for Patch Distribution

Information carried by the ensemble of image patches is noisy. Moreover, patches present varying characteristics between different regions of the face image. We propose to obtain a summary of these noisy and space-varying patches using the joint distribution of these patches modeled by an HMM.

The HMM is a Bayesian network, assuming Markov dependence between hidden states. Training an HMM essentially performs unsupervised segmentation by assigning patches to different hidden states. The distribution of patches assigned to each hidden state is described by a Gaussian mixture density function.

Modelling each face image using a hidden Markov model independently will impose two problems. The number of patches from one image is too limited to robostly estimate an HMM from scratch. Moreover, there would be no correspondence between the Gaussian components in the state-dependent patch distributions (GMMs) in one HMM and the counterparts in another, and neither can we guarantee correspondence between hidden states from different HMMs. As will be shown in Section 4, both correspondences are necessary for constructing the HMM supervector space for face images.

Therefore, we first train a global HMM, which is then adapted to image-specific HMMs for each image by maximum a posterior criterion. The global HMM is trained on all face images, and learns the smoothed common characteristics among the majority of face images, including the transition between different regions of the face image and the initial Gaussian components used in approximating the patch distribution in each region. We derive the image-specific patch-based HMM by adapting the mean vectors within each hidden state of the global HMM.

4 Images in HMM Supervector Space

Critical to the success of face age estimation is to find a human face space less vulnerable to noise and tailored for the space-varying characteristic of human faces. In this section, we propose to construct a human face space using hidden Markov model supervctors and map face images into this space.

Measuring the difference between two face images becomes calculating the difference between the joint distributions of the patches from these two images. Kullback-Leibler divergence is a natural distribution similarity measure, and applies to this problem as in Equation 1.

$$D(f^u||f^v) = \int f^u \log \frac{f^u}{f^v} \tag{1}$$

where f^u and f^v are the distributions of patches for images u and v respectively. When f is modeled by an HMM, a_{ij} denotes the transition probability from hidden state i to hidden state j, and g_i denotes the observation distribution for the i^{th} hidden state.

No closed form solution for calculating the KL divergence between two HMMs is available. However, there exists an upper bound [13] for left-to-right HMMs,

$$\begin{aligned} D(f^u||f^v) \quad \leq \quad & \sum_{i=1}^{S} \{ \frac{D(g_i^u||g_i^v)}{1-a_{ii}^u} + \frac{D(g_i^v||g_i^u)}{1-a_{ii}^v} \\ & + \frac{(a_{ii}^u - a_{ii}^v) log(a_{ii}^u / a_{jj}^v)}{(1-a_{ii}^u)(1-a_{ii}^v)} \} \end{aligned} \tag{2}$$

Since all HMMs in this work share the identical transition probabilities, i.e. those in the global HMM, Equation 2 can be simplified as

$$D(f^u||f^v) \leq \sum_{i=1}^{S} \left\{ \frac{D(g_i^u||g_i^v)}{1-a_{ii}} + \frac{D(g_i^v||g_i^u)}{1-a_{ii}} \right\} \tag{3}$$

Equation 3 shows that we can approximate the difference between two face images using the summation of the Kullback-Leibler divergence between corresponding state-dependent observation distributions in the two HMMs. In our case, the state-dependent observation distribution is described by a Gaussian mixture density function $g(z;\Theta) = \sum_{k=1}^{K} w_k \mathcal{N}(z;\mu_k,\Sigma_k)$, The Kullback-Leibler divergence between two Gaussian mixture density functions has an upper bound from the log-sum inequality,

$$D(g^m||g^n) \leq \sum_{k=1}^{K} w_k D(\mathcal{N}(z;\mu_k^m,\Sigma_k)||\mathcal{N}(z;\mu_k^n,\Sigma_k)), \quad (4)$$

where μ_k^m denotes the adapted mean of the kth component from a Gaussian mixture density g^m, and likewise for μ_k^n.

Since we adapted image-specific HMMs from the global HMM by only updating the component means in Section 3, the covaraiance matrices are identical between two corresponding Gaussian components of the corresponding hidden states. Therefore, the right hand side of Equation 4 is equal to

$$d(g^m,g^n) = \frac{1}{2}\sum_{k=1}^{K} w_k(\mu_k^m-\mu_k^n)^T\Sigma_k^{-1}(\mu_k^m-\mu_k^n). \quad (5)$$

Putting Equations 3 and 5 together, the distance measure between two HMMs is

$$d(f^u,f^v) = \sum_{i=1}^{S}\{\frac{1}{1-a_{ii}}\sum_{k=1}^{K} w_{ik}(\mu_{ik}^u-\mu_{ik}^v)^T\Sigma_{ik}^{-1}(\mu_{ik}^u-\mu_{ik}^v)\}$$

$d(f^u,f^v)$ can be proved to be a metric function, and it can be shown that $d(f^u,f^k) = \left\|\phi(f^u)-\phi(f^k))\right\|$,
where

$$\phi(f^u) = [\sqrt{\tfrac{w_{11}}{1-a_{11}}}\Sigma_{11}^{-\frac{1}{2}}\mu_{11}^u;\cdots;\sqrt{\tfrac{w_{SK}}{1-a_{SS}}}\Sigma_{SK}^{-\frac{1}{2}}\mu_{SK}^u],$$

Each image f is transformed to $\phi(f)$, which we call a patch-based HMM supervector. The similarity between two images is simply the Euclidean distance of the corresponding supervectors.

5. Experiments

We present experiments using patch-based HMM supervector for face age estimation. In this work, face age estimation is approximated as a classification problem, although further performance improvement might be achieved by using the same image representation and distance metric in a regression framework. Each test image is assigned a class label, corresponding to the estimated integer age, i.e., the granularity being one year. Within each age class, there is much variance owing to demographics and cosmetics, among others. Therefore, we adopt Within Class Covariance Normalization (WCCN) [6] to reinforce that in each age class the HMM supervectors are close to each other. The centroids of the normalized HMM supervectors for each age class are estimated. Then we use a nearest centroid classifier, i.e. classifying each test image by assigning the age class label corresponding to the age class with the nearest centroid.

We compare the performance of patch-based HMM supervector with that of GMM supervector. For both approaches, each face image is downsampled to 32 x 32 pixels. 6 x 6 patches are extracted on a dense even grid of 1 x 1 pixels from each image. Each patch is first normalized using its mean and variance. Then we concatenate the 2-D Discrete Cosine Transform of the normalized patch and the x-y coordinates into a 38 dimension vector, as suggested in [7]. The HMM supervector approach adopt 1-D four hidden state left-to-right HMMs with the patches arranged according to their sequence along the horizontal scan lines, switching directions for each line. The patch-based GMM supervector approach is essentially a special case of the proposed HMM supervector with one hidden state. The two approaches both have a total of 512 Gaussian components.

5.1 Datasets and Experiment Setup

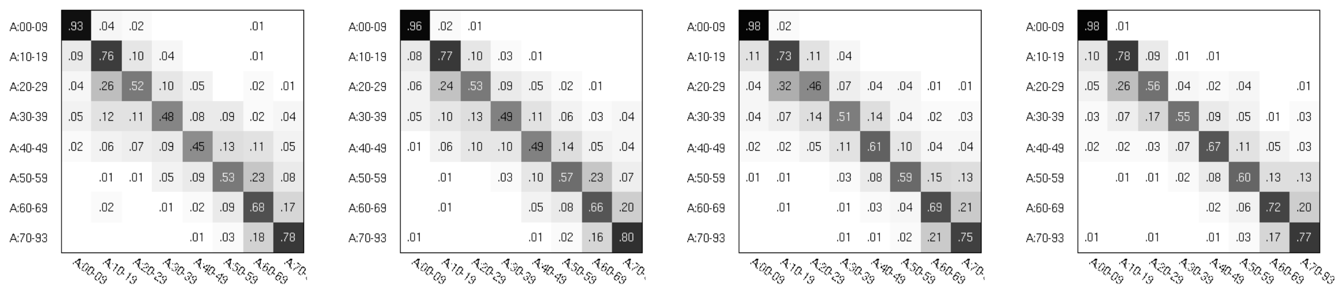
We use a face age dataset containing two subsets of facial images, one with 4000 images of 800 females, and the other with 4000 images of 800 males. Each subset has the age range of 0 to 93. The experiments in this work are carried out separately on female and male subsets. We randomly select half of the images for model training and use the remaining for testing.

We used two measures to evaluate age estimation performance. The first one is Mean Absolute Error (MAE), defined as the average of the absolute errors between the estimated labels and ground truth labels. The second measure is the cumulative score [11] defined as: $CumScore(\theta) = N_{e\leq\theta}/N_t \times 100\%$, where $N_{e\leq\theta}$ is the number of samples on which the absolute errors are not higher than θ.

5.2 Experiment Results

Table 1 presents the MAEs for different age ranges and the average MAEs. Figure 2 presents the cumulative scores at various error levels. The confusion matrices in Figure 3 show the performance if we would just classify the faces into the 10-yr age ranges (with the last range being 70-93).

We can observe that the performance of the proposed patch-based HMM supervector approach compares favorably with the patch-based GMM supervector approach in the average MAEs, in the cumulative scores at almost all error levels and in almost all diagonal elements of the confusion matrices.



(a) GMM Female (b) HMM Female (c) GMM Male (d) HMM Male

Figure 3. Confusion matrices of classifying faces into age ranges.

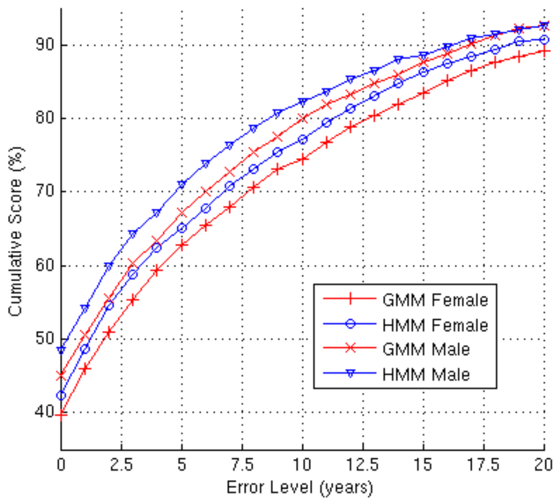


Figure 2. Cumulative scores of age estimation at error levels from 0 to 20 years.

6 Discussion & Conclusion

This study proposes patch-based HMM supervector for face age estimation and improves patch-based GMM supervector approach by tackling three intrinsic problems: 1) loss of spatial structure 2) assumption of identical face patch distribution and 3) lack of correspondence between different regions of faces. Our experiments show that the proposed patch-based HMM supervector performs consistently better than the GMM supervector.

The hidden states in the proposed system could also be used to encode semantic meanings other than the rough spatial structure in this study. For example, [2] found that geometrical structure of different parts of human faces contributes to age estimation. It might be feasible to design a particular patch-based HMM supervector to capture such geometrical structure.

7 Acknowledgements

This work was supported in part by NSF Grant IIS-0534106 and in part by U.S. Government VACE Program.

Table 1. MAEs (year) in different age ranges.

Female			Male		
Range	GMM	HMM	Range	GMM	HMM
0-9	2.9843	1.5866	0-9	1.1457	1.5472
10-19	3.6426	3.5743	10-19	4.3052	3.1325
20-29	7.3846	6.7814	20-29	6.8785	5.8543
30-39	11.492	9.544	30-39	8.572	8.004
40-49	12.729	10.824	40-49	8.2902	7.3725
50-59	8.4106	7.4411	50-59	7.7985	7.1559
60-69	6.0979	6.4766	60-69	5.7362	5.1021
70-93	4.5353	4.1	70-93	3.6316	4.2895
Average	7.246	**6.3335**	Average	5.853	**5.397**

References

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

[2] 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.

[3] 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), pp. 621-628, 200.

[4] X. Geng, Z. Zhou, Y. Zhng, G. Li, and H. Dai. Learning From facial aging patterns for automatic age estimation. *Proceedings of ACMMultimedia*, pp.307-316, 2006.

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

[6] A. O. Hatch, S. Kajarekar, and A. Stolcke Within-class Covariance Normalization for SVM-based Speaker Recognition *ICSLP*, pp. 1471-1474, 2006.

[7] Shuicheng Yan, Ming Liu, Thomas S. Huang Extracting Age Information from Local Spatially Flexible Patches *ICASSP*, 2008

[8] S. Yan, H. Wang, T. Huang, and X. Tang. Auto-Structured Regressor from Uncertain Labels. *ICCV*, 2007.

[9] S. Lucey and T. Chen. A GMM Parts Based Face Representation for Improved Verification through Relevance Adaptation. *CVPR*, pp. 855-861, 2004.

[10] S. Lucey and T. Chen, Learning Patch Dependencies for Improved Pose Mismatched Face Verification. *CVPR*, vol. 1, pp. 909-915, 2006.

[11] X. Geng, Z. Zhou, Y. Zhng, G. Li, and H. Dai. Learning from facial aging patterns for automatic age estimation. *ACM MM*, pp. 307–316, 2006.

[12] Shuicheng Yan, Xi Zhou, Mark Hasegawa-Johnson, and Thomas S. Huang Regression from Patch-Kernel. accepted to *CVPR*, 2008.

[13] Peng Liu, and Frank K. Soong, Kullback-Leibler Divergence between Two Hidden Markov Models. Microsoft Research Asia, Technical Report, 2005.