AGE ESTIMATION FROM HUMAN BODY IMAGES

Yongxin Ge[1,2]*, Jiwen Lu*[3]*, Wu Fan*[1]*, and Dan Yang*[1]

[1]School of Software Engineering, Chongqing University, China
[2]Key Laboratory of Dependable Service Computing in Cyber Physical Society
Ministry of Education, Chongqing University, China
[3]Advanced Digital Sciences Center, Singapore

ABSTRACT

In this paper, we investigate the problem of estimating human ages from full body images. To our best knowledge, this problem has not been formally addressed before possibly due to the great challenges and lacking of such publicly available datasets. However, estimating human ages at a distance has a number of potential applications, especially for visual surveillance in such places as supermarkets, airports, building entrances, and shopping malls. In this paper, we propose a new human age estimation approach from full body images with frontal or back views. Our contributions are three-fold. First, we collect a human body image dataset containing 1500 public figures or celebrities searched from the internet, as well as the age label information of each image. Second, we explore several widely used human local appearance feature descriptors with a regression model to estimate human ages from these body images. Lastly, we apply a multiview canonical correlation analysis (MCCA) method by making use of multiple feature descriptors to exploit complementary information to further improve the age estimation performance. Experimental results have clearly demonstrated the feasibility of using fully body images to estimate human age and the efficacy of our proposed approach.

Index Terms— Age estimation, multiple feature fusion, biometrics.

Fig. 1. Examples of human body images of ten different subjects in our dataset, where the top and bottom rows are body images for female and male, and the number below each image is the age value of the person, respectively. The objective of this study is to learn an age estimator to predict the age information of the person from his/her body image.

1. INTRODUCTION

Human age estimation has become a particularly interesting research topic in computer vision and biometrics in recent years due to its potential applications such as intelligent security control, visual surveillance, biometrics authentication and social media analysis [1, 2, 3, 4, 5]. Existing human age estimation methods mainly focus on facial age estimation, and a number of such algorithms have been proposed in the literature [1, 2, 3, 4, 6, 7, 8]. While many existing facial age estimation methods have attained reasonably good performance when the training and test face data are acquired under similar conditions, they require face images of the subjects concerned be acquired in high-resolution such that more discriminative information can be used to learn the age estimator. In some practical applications such as visual surveillance, there is usually some distance between the camera and the persons of interest, and facial images collected under such scenarios are usually of low-resolution or with limited details. To address this, some researchers have utilized human gait signatures for age estimation in recent years [9, 10, 11]. However, extraction of human gait signatures needs a relatively complete video clip, which is very hard to obtain in many real applications.

In this paper, we investigate the problem of age estimation from human body images. To our best knowledge, this problem has not been formally addressed before possibly due to the great challenges and lacking of such publicly available datasets. However, estimating human ages at a distance has a number of potential applications, especially for visual surveillance in such places as supermarkets, airports, building entrances, and shopping malls. In such scenarios, both face im-

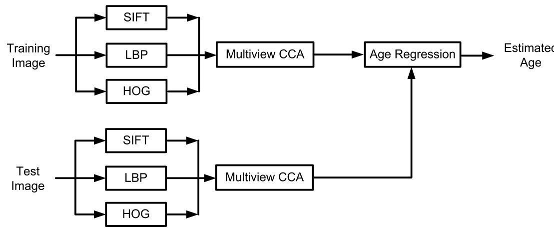

Fig. 2. Flow-chart of our proposed human age estimation approach. For each image in the training set, we extract three different local features: HOG, LBP, and SIFT. Then, we fuse these features into a common low-dimensional feature subspace by using multiview CCA. Based on the fused features, we learn an age regressor to model the relationship of the human body image and the age value. For a test image, we also extract three local features and use multiview CCA to combine these features into a feature vector. Lastly, we apply the learned age regressor to predict the age value of the testing human body image.

ages and gait sequences are not stable for human age estimation. However, human body image can be easily obtained and recent advances in computer vision have also shown that human body is an effective cue for soft biometrics, such as gender classification from human body images [12, 13, 14, 15]. Motivated by these findings, we propose in this paper a new human age estimation approach from full body images with frontal or back views. Our contributions are three-fold. First, we collect a human body image dataset containing 1500 public figures or celebrities searched from the internet, as well as the age label information of each image. Second, we explore several widely used human local appearance feature descriptors with a regression model to estimate human ages from these body images. Lastly, we apply a multiview canonical correlation analysis (MCCA) method by making use of multiple feature descriptors to exploit complementary information to further improve the age estimation performance. Experimental results have clearly demonstrated the feasibility of using fully body images to estimate human age and the efficacy of our proposed approach.

2. PROPOSED APPROACH

Fig. 2 shows the flow-chart of our proposed approach to age estimation from human body images.

2.1. Feature Representation

Raw pixel is the most straightforward representation for human body images. However, it may be not a good choice for our body-based human age estimation task because it will suffers from large variations of human clothes, poses, illumination, occlusion in practical applications. To address this, we adapt the following three local feature representation methods which have been successfully used in many image analysis tasks due to their strong robustness:

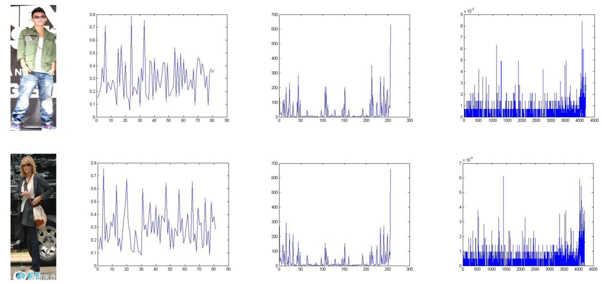

Fig. 3. Two samples of our human body image database and the corresponding local feature representations. For left to right in each row are the original raw images, HOG, LBP and SIFT local feature representations, respectively.

Histogram of Oriented Gradient (HOG) [16]: HOG represents edges with a magnitude-weighted histogram and groups these information according to the edge direction. Following [16], we divided each body image into 3×3 cell blocks of 6×6 pixel cells, and represented each block by a HOG feature, as a 9-dimensional feature vector describing the gradient information from 9 different orientations. HOG was originally designed for human detection, and has been recently applied to face recognition [17] and gender recognition [12]. The reason we applied it to our age estimation task is we expect this feature to extract the shape information of human body images.

Local Binary Pattern (LBP) [18]: LBP feature descriptor was originally designed for texture description [18], and then has been widely used in many other visual analysis tasks such as face recognition [19], facial expression recognition [20], and human detection [21, 22]. The basic idea of LBP is to assign a label to every pixel of a gray-scale image by thresholding the neighborhoods of this pixel with the center pixel value and considering the result as a binary number. Then, the histogram of these labels are used as a texture descriptor. The motivation of using LBP for human body description is inspired by the fact that human body can be considered as a composition of micro-patterns which can be well described by this operator. Hence, we expect this feature descriptor can extract the micro-patterns of human body shape ro reflect the age information.

Scale-Invariant Feature Transform (SIFT) [23]: SIFT is a very powerful feature descriptor and has been successfully used in numerous computer vision applications [23]. Since the SIFT feature can gain invariance to scale and rotation by exploiting scale-space extrema and the local dominant orientation, it has demonstrated excellent discriminative power in object recognition. To better utilize the spatial information, this feature can be extracted in a spatial pyramid manner [24], where the original image is divided into 21×21 segments from different scales and the histogram of SIFT feature within each segment was extracted and concatenated.

Fig. 3 shows two examples of human body image and the corresponding local feature representations.

2.2. Multiview CCA

Let $X = [x_1, x_2, \cdots, x_N]$, $Y = [y_1, y_2, \cdots, y_N]$, and $Z = [z_1, z_2, \cdots, z_N]$ be the feature sets extracted from the HOG, LBP and SIFT feature descriptors, respectively, where N is the number of samples in the training set. Multiview CCA aims to seek three projection vectors, w_x, w_y, and w_z such that the joint pairwise correlations between $x = w_x^T x$, $y = w_y^T y$ and $z = w_z^T z$ is maximized [25]:

$$\begin{aligned} \max_{w_x, w_y, w_z} \quad & w_x^T XY^T w_y + w_x^T XZ^T w_z + w_y^T YZ^T w_z \\ \text{subject to:} \quad & w_x^T B_x w_x + \alpha w_y^T B_y w_y + \beta w_z^T B_z w_z = 1 \end{aligned} \tag{1}$$

where $B_x = \frac{1}{N} XX^T$, $B_y = \frac{1}{N} YY^T$, and $B_z = \frac{1}{N} ZZ^T$ denote the covariance matrices of the HOG, LBP and SIFT feature descriptors, respectively, and α and β are two parameters to balance the contributions of different features, and they were empirically set as 1 and 1, respectively.

The expression in Eq. (1) can be simplified to the following form:

$$\begin{aligned} \begin{bmatrix} w_x^* \\ w_y^* \\ w_x^* \end{bmatrix} &= \arg \max_{w_x, w_y, w_z} \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix}^T A \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} \\ \text{subject to} \quad & \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix}^T B \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} = 1. \end{aligned} \tag{2}$$

where

$$A = \begin{bmatrix} 0 & \frac{1}{2} w_x^T XY^T w_y & \frac{1}{2} w_x^T XZ^T w_z \\ \frac{1}{2} w_x^T XY^T w_y & 0 & \frac{1}{2} w_y^T YZ^T w_z \\ \frac{1}{2} w_x^T XZ^T w_z & \frac{1}{2} w_y^T YZ^T w_z & 0 \end{bmatrix} \tag{3}$$

$$B = \begin{bmatrix} B_x & 0 & 0 \\ 0 & \alpha B_y & 0 \\ 0 & 0 & \beta B_z \end{bmatrix} \tag{4}$$

Let $w = \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix}$. Eq. (1) can be further rewritten as

$$\begin{aligned} w^* &= \arg \max w^T A w \\ \text{subject to} \quad & w^T B w = 1. \end{aligned} \tag{5}$$

Lastly, the projections of multiview CCA can be easily obtained by solving the following generalized eigenvalue equation:

$$Aw = \lambda B w. \tag{6}$$

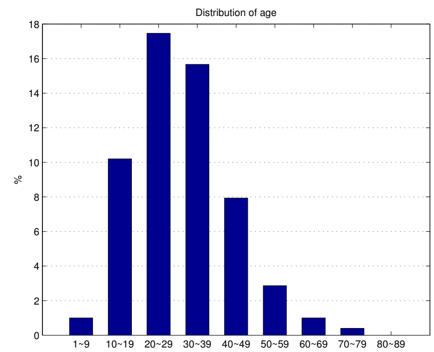

Fig. 4. Age distribution of our dataset.

Let $w_1, w_2, \cdots, w_k$ be the eigenvectors of Eq. (6) corresponding to the k largest eigenvalues ordered according to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$. A transformation matrix $W = [w_1, w^2, \cdots, w^k]$ is the projection matrix of the multiview CCA method.

Having obtained the projection vectors w_x, w_y and w_z, we use the linear regression method [26] to model the multiview features and the age values.

3. EXPERIMENTS

In this section, we evaluate our proposed approach by conducting age estimation from human body images experiments on our dataset. The following describes the details of the experiment setups and results.

3.1. Dataset

In order to make a good study on the problem of age estimation from human body images, a large database of human body images with age labels is required. To the best of our knowledge, however, there is no such database publicly available. To this end, we collected 1500 human body images by searching from the internet, which are all public figures or celebrities due to easily acquiring their age information. We impose no restrictions of scale, pose, lighting, background, ethnicity and partial occlusion, such that a fully automatic and real-time system which takes general human body images as inputs can be achieved, which has great potential in real-word applicants. Our dataset contains 1500 images across three different ethnicity: Asian, Caucasian and African. There are 848 male and 652 female images in the dataset, respectively. The age range of our dataset is [8, 83]. Fig. 4 shows the age distribution of our dataset, and some example human body images can be seen in Fig. 1.

Table 1. MAE (years old) and standard deviation comparison of different local features.

Method	MAE	standard deviation
HOG	9.44	0.46
LBP	11.86	0.38
SIFT	9.15	0.27

Table 2. MAE (years old) and standard deviation comparison of different multiple feature based age estimation methods.

Method	MAE	standard deviation
Feature concatenation	9.80	0.52
Our approach	8.80	0.25

3.2. Experimental Settings

The quantitative metrics of evaluations are very essential to demonstrate the performance of age estimation. We use the Mean Absolute Error (MAE) to evaluate the performance of our method, defined as:

$$MAE = \frac{1}{N_T}\sum_{k=1}^{N_T}|\hat{l_k} - l_k| \tag{7}$$

where $\hat{l_k}$ and l_k are the estimated and label of the kth testing image sample, and N_T is the number of testing body image samples.

We perform all processing in grayscale, and each experiment is repeated ten times with randomly selected 1000 images to construct the training set and the remaining to form the testing set. The final result is reported as the mean and standard deviation of the results from the ten runs.

3.3. Results and Analysis

Single Feature Performance: We first investigated the discriminative power of different local feature representation methods. For all these local features, we applied PCA to project them into 60-dimensional subspace. Table 1 shows the mean and standard deviation of MAEs for these three feature representations. We can see that SIFT achieves the best performance and LBP performs worst. Specifically, SIFT outperforms HOG and LBP with 3.1% and 22.3% improvement in terms of the MAE, respectively. Moreover, we can also observe from this table that SIFT demonstrates the best robustness while HOG shows the worst.

Multiple Feature Performance: We compared our method with the feature concatenation method in our age estimation task. Specifically, we concatenated three features of each body image into a long feature and then performed age estimation. Table 2 shows the mean and standard deviation of MAEs for different multiple feature based age estimation methods. We can see that our method achieves better performance than the feature concatenation method.

Table 3. MAE (years old) comparisons of existing age estimation methods based on human face and gait.

Method	Cue	MAE	Dataset
Method in [27]	Human face	9.79	4000 subjects
Method in [2]	Human face	3.91	4000 subjects
Method in [9]	Human gait	6.23	122 subjects
Method in [11]	Human gait	8.20	1728 subjects
Our method	Human body	8.80	1500 subjects

Table 4. MAE (years old) comparison of human ability and our approach on age estimation from human body images.

Method	MAE
Human	7.44
Our method	8.80

Comparison with Existing Age Estimation Approaches: We compared our age estimation approach with state-of-the-art facial age estimation and gait-based human age estimation methods. Since different features and datasets were adopted, it is not easy to compare them under the same settings. The aim of this comparison is to give some conceptual analysis for human age estimation with different cues. Table 3 lists some results, and we can observe from this table that our method can achieve comparable performance with these age estimation methods based on human face and gait.

Comparison with Human Observers: As an important baseline, the human ability in age estimation from human body images is also tested. We randomly selected 100 images aged from $[15, 60]$ from our dataset to 10 human observers (7 males and 3 females) with age of 20 to 30 years old. None of them received training on the task before the experiment. Table 4 shows the MAE of human ability on age estimation on our dataset. We can observe from this table that our proposed automatic age estimation approach performs slightly worse than human observers.

4. CONCLUSION

We have investigated the problem of estimating human ages from full body images. By applying three different feature representation methods to describe human body images and fusing them by multiview CCA, an automatic age estimation framework from human body images is proposed. Experimental results have clearly demonstrated the feasibility of using fully body images to estimate human age and the efficacy of our proposed approach.

5. ACKNOWLEDGEMENT

This work is supported by the Fundamental Research Funds for the Central Universities (CDJZR12090002).

References

[1] Y. H. Kwon and N. V. Loba, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.

[2] G. Guo, Y. Fu, CR Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

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

[4] A. Lanitis, C. J. Taylor, and T.F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.

[5] S. Yan, H. Wang, X. Tang, J. Liu, and T. S. Huang, "Regression from uncertain labels and its applications to soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 698–708, 2008.

[6] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifld analysis of face pictures and regression on age features," in *IEEE International Conference on Multimedia and Expo*, 2007, pp. 1383–1386.

[7] J. Lu and Y.P. Tan, "Fusing shape and texture information for facial age estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 1477–1480.

[8] J. Lu and Y.P. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 2, pp. 249–258, 2013.

[9] J. Lu and Y.-P. Tan, "Gait-based human age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 761–770, 2010.

[10] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, 2012.

[11] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in *International Joint Conference on Biometrics*, 2011, pp. 1–6.

[12] L. Cao, M. Dikmen, Y. Fu, and T.S. Huang, "Gender recognition from body," in *ACM International Conference on Multimedia*, 2008, pp. 725–728.

[13] M. Collins, J. Zhang, P. Miller, H. Wang, and H. Zhou, "Eigenbody: Analysis of body shape for gender from noisy images," in *International Machine Vision and Image Processing Conference*, 2010.

[14] G. Guo, G. Mu, and Y. Fu, "Gender from body: A biologically-inspired approach with manifold learning," in *ACCV*, 2010, pp. 236–245.

[15] M. Collins, J. Zhang, P. Miller, and H. Wang, "Full body image feature representations for gender profiling," in *IEEE International Conference on Computer Vision Workshops*, 2009, pp. 1235–1242.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE International Coneference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[17] A. Albiol, D. Monzo, A. Martin, Sastre J., and A. Albiol, "Face recognition using hog-ebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.

[18] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[19] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[20] C. Shan, S. Gong, and P.W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[21] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative local binary patterns for human detection in personal album," in *IEEE Interntaional Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[22] X. Wang, T.X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *IEEE 12th International Conference on Computer Vision*, 2009, pp. 32–39.

[23] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[24] N. Dalal and B. Triggs, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *IEEE International Coneference on Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.

[25] A. Sharma, A. Kumar, H. Daume, and D.W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2160–2167.

[26] Y. Fu, S. Yan, and T.S. Huang, "Classification and feature extraction by simplexization," *IEEE Transactions Information Forensics and Security*, vol. 3, no. 1, pp. 91–100, 2008.

[27] S Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.