Expression-Invariant Age Estimation

Fares Alnajar*[,1]
F.alnajar@uva.nl
Zhongyu Lou*[,1]
z.lou@uva.nl
Jose Alvarez[2]
jose.alvarez@nicta.com.au
Theo Gevers[1]
th.gevers@uva.nl

[1] ISLA Lab, Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
[2] NICTA
Canberra ACT 2601
Australia

We investigate and exploit the influence of facial expressions on automatic age estimation. Different from existing approaches, our method jointly learns the age and the expression by introducing a new graphical model with a latent layer between the age/expression labels and the features. This layer aims to learn the relationship between the age and the expression and captures the face changes which induce the aging and the expression appearance, and thus obtaining expression-invariant age estimation.

External factors like facial expressions cause changes in facial muscles which distort the aging cues. A problem in age estimation is that expression-related muscles overlap with aging-induced facial changes. For example, smiling involves the activation of some facial muscles leading to raising the cheeks and pulling the lip corners. This influences the aging wrinkles around the mouth and near the eyes. Consequently, the aging cues changes caused by expressions show the necessity of separating the influence of expression when estimating the age.

We jointly learn the age and expression and model their relationship. More specifically, we introduce a new graphical model which contains a latent layer between the age/expression labels and the facial features. This layer captures the relationship between the age and expression. To predict the age, the age and expression are inferred jointly, and hence prior-knowledge of the expression of the test face is not required. The contributions of our work are: 1) we show how age-expression joint learning improves the age prediction compared to learning independently from expression. 2) As opposed to existing methods [2, 5], the proposed method predicts the age across different facial expressions without prior-knowledge of the expression labels of the test faces. 3) Finally, our results outperform the best reported results on age-expression datasets (FACES [1] and Lifespan [3]).

The proposed graphical model has four sets of connections: First, connections between the face subregions and the latent variables. These connections are designed to capture the changes of face appearance related to age and expression. Second, connections between the face subregions and the age/expression labels are formed. The aim here is to directly infer the age/expression from the features. Third, connections between the latent variable modeling the relationship between the face subregions. Finally, connections are established between the latent variables, the age, and the expression. The last type of connections is designed to relate the age with the expression which allows the joint learning between them.

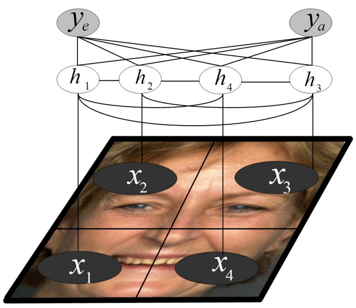

Figure 1: Our graphical model to jointly learn the age and the expression. $\mathbf{x} = [x_1, x_2, x_3, x_4]$ represents the feature vector, $\mathbf{h} = [h_1, h_2, h_3, h_4]$ denotes the latent variables, y_a and y_e are the corresponding age and expression respectively. Note that, while all x_i are connected with y_a and y_e, we do not show these connections in this figure for the sake of clarity.

Our model maximizes the conditional probability of the joint assignment of $\mathbf{y}$ given observation $\mathbf{x}$:

$$\mathbf{y}^* = argmax_{\mathbf{y}} P(\mathbf{y}|\mathbf{x};\theta). \tag{1}$$

$$P(\mathbf{y}|\mathbf{x};\theta) = \sum_{\mathbf{h}\in\mathcal{H}} P(\mathbf{y},\mathbf{h}|\mathbf{x};\theta) = \frac{\sum_{\mathbf{h}\in\mathcal{H}} exp(\psi(\mathbf{y},\mathbf{h},\mathbf{x};\theta))}{\sum_{\mathbf{y}'\in\mathcal{Y},\mathbf{h}\in\mathcal{H}} exp(\psi(\mathbf{y}',\mathbf{h},\mathbf{x};\theta))}.$$

Where $\psi(.)$ is the potential function which measures the compatibility between the (observed) features, the joint assignment of the latent variables, and the output labels. The potential function is decomposed into four potential functions corresponding to the connections of the model (Figure 1).

$$\psi(\mathbf{y},\mathbf{h},\mathbf{x};\theta) = \sum_{i=1}^{4} \psi_1(y_a,x_i;\theta_i^1) + \sum_{i=1}^{4} \psi_2(y_e,x_i;\theta_i^2) + \sum_{i=1}^{4} \psi_3(h_i,x_i;\theta_i^3) + \psi_4(\mathbf{h},y_a,y_e;\theta^4). \tag{2}$$

To learn the parameters θ, we exploit the max margin approach [4]. The inference involves a combinatorial search of the joint assignment of $\mathbf{h}$, y_e and y_a which results in the maximum conditional probability:

$$(\hat{\mathbf{y}},\hat{\mathbf{h}}) = argmax_{\mathbf{y}\in\mathcal{Y},\mathbf{h}\in\mathcal{H}} \psi(\mathbf{x},\mathbf{y},\mathbf{h};\theta). \tag{3}$$

In the paper, we evaluate our model on FACES [1] (6 expressions) and Lifespan [3] (2 expressions) datasets. The experiments show the improvement in performance when the age is jointly learnt with the expression in comparison to expression-independent age estimation. The age estimation error is reduced by 14.43% and 37.75% for FACES and Lifespan datasets respectively. We show (Figure 2) the face regions corresponding to each hidden state (3).

Figure 2: Average face regions corresponding to different hidden states (from left to right) for the bottom and top face regions. For the bottom regions, the first hidden state corresponds to the face appearance where the mouth is open, the third hidden state represents a depressed lip corner, and the second hidden state corresponds to a normal face appearance. For the top regions, the second hidden state represents the face appearance where the eye is slightly closed while the first and the third states correspond to open eye appearances.

[1] Natalie C. Ebner, Michaela Riediger, and Ulman Lindenberger. Faces: database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1):351–362, 2010.

[2] Guodong Guo and Xiaolong Wang. A study on human age estimation under facial expression changes. *Computer Vision and Pattern Recognition*, pages 2547–2553, 2012.

[3] Meredith Minear and Denise C. Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, and Computers*, 36(4):630–633, 2004.

[4] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, pages 1453–1484, 2005.

[5] Chao Zhang and Guodong Guo. Age estimation with expression changes using multiple aging subspaces. *Biometrics: Theory, Applications and Systems*, pages 1–6, 2013.

* Both authors contributed equally.