UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Expression-Invariant Age Estimation

Alnajar, F.; Lou, Z.; Alvarez, J.; Gevers, T.

DOI
10.5244/C.28.14

Publication date
2014

Document Version
Final published version

Published in
Proceedings of the British Machine Vision Conference 2014

Link to publication

Citation for published version (APA):
Alnajar, F., Lou, Z., Alvarez, J., & Gevers, T. (2014). Expression-Invariant Age Estimation. In M. Valstar, A. French, & T. Pridmore (Eds.), *Proceedings of the British Machine Vision Conference 2014* (pp. 14.1-14.11). BMVA Press. https://doi.org/10.5244/C.28.14

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:14 Nov 2026

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

Abstract

In this paper, 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.

Conducted on two age-expression datasets (FACES [9] and Lifespan [10]), our experiments illustrate the improvement in performance when the age is jointly learnt with expression in comparison to expression-independent age estimation. The age estimation error is reduced by 14.43% and 37.75% for the FACES and Lifespan datasets respectively. Furthermore, the results obtained by our graphical model, without prior-knowledge of the expressions of the tested faces, are better that the best reported ones for both datasets.

1 Introduction

Automatic age estimation is an important research field in the area of computer vision and has many applications such as human-computer interaction, security, and surveillance. In general, the human age is derived from facial aging cues. The aging of adults is primarily perceived via skin changes [6]. During aging, the human face loses collagen beneath the skin leading to thinner, darker, and more leathery skin [6]. Age-induced facial wrinkles become more distinct as a result of repeated activation of facial muscles and they start to appear in different directions depending on these muscles [7]. For example, vertical wrinkles intensify between the eyebrows while horizontal wrinkles become more apparent close to the eye corners.

External factors like facial expressions cause changes in facial muscles which distort the aging cues. A facial expression is explained by a combination of these changes in the face which are called Action Units [5]. A problem in age estimation is that expression-related muscles overlap with aging-induced facial changes. For example, smiling involves

© 2014. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.
*Both authors contributed equally.

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.

Most of the existing age estimation methods assume that faces show little or no expressions and ignore the changes of the face appearance induced by them. Guo et al. [7] study human age estimation under facial expression changes. Their method learns the correlation between two expressions at a time (e.g. neutrality and happiness). To predict the age across two expressions, the face is mapped from one expression (e.g. happiness) to another (e.g. neutrality). Next, the age is predicted from the "mapped" face. For the face aging representation, Biologically-Inspired Features (BIF) [8] and Marginal Fisher Analysis (MFA) are used. Zhang et al. [18] employ a weighted random subspace method to solve cross-expression age estimation. In their method, several feature sets are generated first, then subspaces are built for these sets. Next, a classifier is learnt for each subspace and predictions of all classifiers are fused to produce the final prediction. Their method does not require different expressions from the same subjects as opposed to [7]. However, both methods [7, 18] require the expressions of test images to be known before predicting the age which limits their applicability.

In this paper, we propose a different approach. Instead of learning the age across two expressions, we jointly learn the age and expression and model their relationship. The aim is to achieve expression-invariant age estimation. In our approach, one model is learnt for all expressions. 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. 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. During training, the age and expression variables are observed. This allows the latent layer to learn the configurations which map the features to the age for different expressions and thus obtaining expression-invariant age estimation. For testing, the age and expression labels are unknown and the method finds the values of age, expression and latent layer which together maximize their compatibility with the features.

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, 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 and Lifespan).

2 Algorithm

The proposed graphical model aims to jointly learn the relationship between age and expression. To this end, an inter-connected latent layer is introduced. The latent variables encode the changes in face appearance. These variables are not explicitly defined, but learnt from the training data.

The 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. Next, we discuss the model formulation and explain the inference and learning techniques.

2.1 Model Formulation

Suppose we have N training samples (images) $\{\mathbf{s}_1 = (\mathbf{x}_1, \mathbf{y}_1), \mathbf{s}_2 = (\mathbf{x}_2, \mathbf{y}_2), ..., \mathbf{s}_N = (\mathbf{x}_N, \mathbf{y}_N)\}$ where $\mathbf{x}_n$ represents the features for sample $\mathbf{s}_n$ and $\mathbf{y}_n = \{y_{a,n}, y_{e,n}\} \in \mathcal{Y} = \mathcal{A} \times \mathcal{E}$ denotes the age and the expression labels. $\mathcal{A}$ and $\mathcal{E}$ are the age and the expression spaces, respectively. The image is uniformly divided into four (2×2) sub-regions. The feature vector extracted from each sub-region x_i is connected to the corresponding hidden variable h_i. Hence, the sample feature vector consists of four sub-region vectors $\mathbf{x}_n = [x_1, x_2, x_3, x_4]$ and the corresponding latent layer is denoted by $\mathbf{h}_n = [h_1, h_2, h_3, h_4] \in \mathcal{H}^4$, where $\mathcal{H}$ is the space of the latent variable state.

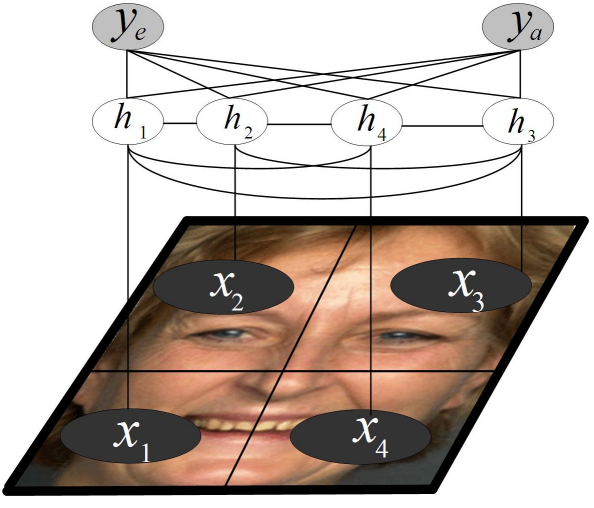

Figure 1: Our graphical model to jointly learn the age and the expression. $\mathbf{x}$ represents the feature vector, $\mathbf{h}$ denotes the latent variables, y_a and y_e are the corresponding age and expression respectively. Note that, while all x_i are connected with y_a and y_e, we do not show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

The aim is to learn the mapping between the features $\mathbf{x}$ and labels $\mathbf{y}$. Our model maximizes the conditional probability of the joint assignment of $\mathbf{y}$ given observation $\mathbf{x}$:

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

Where:

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

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. In the next section, the potentials are defined.

2.2 Potentials

The potentials measure the compatibility of the joint assignment of different variables:

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

In our model, four types of potentials are used. Hereafter, we explain each one of them. Potential ψ_1 models the compatibility of the features and the age:

$$\psi_1(y_a,x_i;\theta_i^1) = \theta_i^1\phi_1(y_a,x_i), \tag{3}$$

where $\phi_1(y_a,x_i)$ represents the feature mapping function encoding the features of the joint assignment of y_a and x_i. The length of $\phi_1(y_a,x_i)$ is equal to the length of x_i multiplied by the cardinality of y_a. In case there are S different ages and the feature vector x_i has K features, the size of θ_i^1 will be $S\times K$. The mapped feature vector is given by:

$$\phi_1(y_a,x_i) = [\underbrace{0\ldots0}_{K\times(y_a-1)dimension}\quad x_i^T\ldots0]. \tag{4}$$

The model turns into a multi-class SVM for age estimation when solely this potential is utilized with the maximum margin method. Multi-class SVM is used as a baseline in this paper. This potential models the global mapping between the input features and the output age prediction.

Potential ψ_2 models the compatibility of the features and the expression:

$$\psi_2(y_e,x_i;\theta_i^2) = \theta_i^2\phi_2(y_e,x_i), \tag{5}$$

where $\phi_2(y_e,x_i)$ encodes the features of the joint assignment of y_e and x_i and is defined in the same way as in equation 4.

Potential ψ_3 models the compatibility of the observation and the latent states:

$$\psi_3(h_i,x_i;\theta_i^3) = \theta_i^3\phi_3(h_i,x_i). \tag{6}$$

Here, $\phi_3(h_i,x_i)$ encodes the features of the joint assignment of the latent variable h_i and the features x_i. The latent variables capture the changes of face appearance. For example, a hidden state could represent whether the mouth is open (e.g. happy) or frowning (e.g. angry). Thus, the potential $\psi_3(h_i,x_i;\theta)$ learns the mapping of the observed features to the appearance changes.

The potential ψ_4 models the compatibility between the age, the expression, and the latent layer:

$$\psi_4(\mathbf{h},y_a,y_e;\theta^4) = \theta^4\phi_4(\mathbf{h},y_e,y_a). \tag{7}$$

$\phi_4(\mathbf{h},y_e,y_a)$ represents the feature mapping function which encodes the features of the joint assignment of $\mathbf{h}$, y_e and y_a. The length of $\phi_4(\mathbf{h},y_e,y_a)$ is the multiplication of the cardinalities of $\mathbf{h}$, y_e and y_a. The element corresponding to the assignment of $\mathbf{h}$, y_e and y_a is set to be 1 while all other elements are set to be 0.

2.3 Inference and Learning

Inference: Given the model parameters θ, 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:

$$(\hat{\mathbf{y}}, \hat{\mathbf{h}}) = argmax_{\mathbf{y}\in\mathcal{Y}, \mathbf{h}\in\mathcal{H}} \psi(\mathbf{x}, \mathbf{y}, \mathbf{h}; \theta). \tag{8}$$

Since the proposed graphical model contains loops, it is intractable in general to perform the maximization. However, by collapsing all the latent variables $\mathbf{h}$ with the output variables y_e a new potential factor is obtained. In the same way, by collapsing all the latent variables $\mathbf{h}$ with the output variables y_a we get another new potential factor. Then the model becomes a chain structure and dynamic method is used to solve the maximization problem [11].

Learning: To learn the parameters θ, we exploit the max margin approach [14]. Since the latent variables $\mathbf{h}$ are not labeled in the training set, we need to solve the following latent structure SVM problem:

$$\min_{\theta,\xi} \left\{ \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{N} \xi_i \right\} \tag{9}$$
$$s.t.\ \forall i \in \{1, 2, \ldots, N\}, \forall \mathbf{y}, \forall \mathbf{h} \in \mathcal{H}:$$
$$\xi_i \geq \Delta(\mathbf{y}_i, \mathbf{y}) + \psi(\mathbf{y}, \mathbf{h}, \mathbf{x}_i; \mathbf{w}) - \psi(\mathbf{y}_i, \mathbf{h}_i^*, \mathbf{x}_i; \theta).$$

Where $\psi(.)$ is the potential function as described in equation 2. $\mathbf{h}_i^*$ is the optimum state under the current parameter. The loss function $\Delta(\mathbf{y}_i, \mathbf{y})$ is defined as the following:

$$\Delta(\mathbf{y}, \hat{\mathbf{y}}) = \begin{cases} |y_a - \hat{y}_a| & if\ y_e = \hat{y}_e \\ 1 + |y_a - \hat{y}_a| & if\ y_e \neq \hat{y}_e \end{cases}. \tag{10}$$

This optimization problem is non-convex. Following [16], we use the CCCP concave-convex framework [17] to solve it. More details of the CCCP procedure can be found in [16, 17].

3 Experiments

The goal of the proposed approach is to capture the relationship between the age and expression and, hence, alleviate the influence of expression in age estimation. In this section, we conduct a number of experiments to validate our model using the age-expression datasets FACES [4] and Lifespan [10].

3.1 Datasets

The publicly available age estimation datasets like FG-NET [1] and MORPH [13] contain mostly neutral faces. The non-neutral faces in those datasets are not expression-labeled. Therefore, to evaluate expression-invariance age estimation, we use other datasets: FACES and Lifespan, which are recently introduced to the computer vision community [7]. FACES dataset contains face images of 171 subjects showing 6 basic expressions: neutrality, happiness, anger, fear, disgust, and sadness. Every subject shows all the expressions resulting in $1026 = 171 \times 6$ face images. The faces in the dataset are frontal with fixed illumination mounted in front and above of the faces. The ages of the subjects range from 19 to 80. The age distribution is not uniform and in total there are 37 different ages.

The Lifespan dataset is a collection of faces of subjects from different ethnicities showing different expressions. The expression subsets have the following sizes: 580, 258, 78, 64, 40, 10, 9, and 7 for neutrality, happiness, surprise, sadness, annoyed, anger, grumpy, and disgust, respectively. The ages of the subjects range from 18 to 93 years and in total there are 74 different ages. The dataset has no labeling for the subject identities. We follow the setup of [7, 18] and use the neutral and the happy subsets. Although the age distributions of both datasets cover a wide range of ages, the FACES dataset is more challenging for age prediction since its expression variation (six expressions) is larger than the one in Lifespan dataset (two expressions).

For feature extraction, eye centers are first automatically detected and the faces are registered and cropped. Then, the faces are divided into 8×8 patches and a local feature vector is extracted for each patch. Finally, the patch local descriptors are concatenated together to form the face descriptor. To extract the features from each patch, we use Local Binary Pattern (LBP) [17]. It is a simple, efficient, and rotation-invariant approach and successfully used for age prediction to capture the skin texture details [7, 15]. In our experiments, we use 8 sampling points with a radius equal to 1.

As in previous setups [7, 18], the datasets are divided into 5 folds. For the FACES dataset, the expression distributions are uniform for all the 5 folds, and none of the subjects appears in more than one fold. For the Lifespan dataset, the dataset (neutral and happy) is split randomly into 5 folds. As the subject identities are not available, a subject overlap between the training and the test samples is possible. The results are measured quantitatively by Mean Absolute Error (MAE) $\frac{1}{N}\sum_{n=1}^{N}|y_a^n - \hat{y}_a^n|$. Where y_a^n is the true age for the test sample n, $\hat{y}_a^n$ is the predicted age for the test sample n, and N is the number of the test samples.

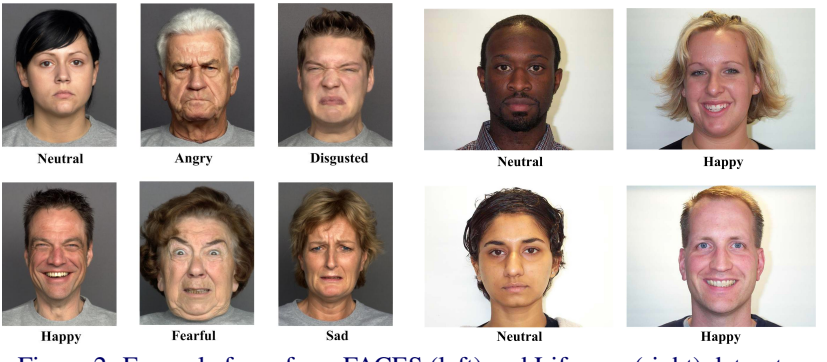

Figure 2: Example faces from FACES (left) and Lifespan (right) datasets.

3.2 Expression-Invariant Age Estimation

In this experiment, we compare two cases. First, learning the age independently from the expression. Second, learning the age jointly with the expression. In both cases, the same 5-fold age-expression datasets are used for evaluation. For the expression-independent learning, a multi-class SVM is used as a baseline. In the expression-joint learning, we use the proposed graphical model and the number of hidden states $|\mathcal{H}|$ is set to 3. For the model learning, the expression is observed and the potential function in equation 2 is applied. The results for the proposed model are shown in Table 1. For both datasets, our graphical model significantly reduces the prediction error (14.43% for FACES and 37.75% for Lifespan). The

errors reported in [7] and [18] for FACES and Lifespan datasets are shown in Table 1. Although both methods assume prior-knowledge of the expression of tested samples, our model outperforms their results for the two datasets.

We further compare our age estimation approach with the joint classification method by [9]. The method was proposed to recognize facial expressions while reducing the influence of human aging. In their method, the authors simply divide the dataset into four age groups ([18-29],[30-49],[50,69], and [70-94]) and consider each expression within each age group as a new class. Then, classification is performed on the newly defined classes. For facial feature extraction, they manually labelled 31 fiducial points and applied Gabor filters [3] on the locations of those points. The four age group classification accuracy using the joint learning method is reported.

To make a fair comparison, and since the authors [9] manually labeled 31 fiducial points on the face, we use our features and compare only the joint learning methods. To this end, we create a new class for each age/expression combination. Different from [9], where the datasets are divided into four age groups, we consider each age separately. The total number of new classes is $37 \times 6 = 222$ in the FACES dataset and $74 \times 2 = 148$ classes in the Lifespan dataset. The obtained errors for FACES and Lifespan are 9.94 and 8.85 years respectively, which are higher than the baseline errors and the ones obtained by our graphical model. It is worth mentioning that in [9], as the datasets are divided into four age groups, the method is tested on smaller number of "joint-classes" (24 and 8 for FACES and Lifespan respectively). In this experiment, the number of joint-classes is much higher.

Detailed results for independent and joint learning for FACES and Lifespan datasets are shown in Tables 2 and 3, where the error for each expression subset is shown separately. The error is reduced for all expression subsets, however, in different rates. The largest improvement is achieved for neutrality (with 30.13% error reduction), while the smallest improvement is obtained for the anger and the disgust expressions (4.64% and 2.43% respectively). This is explained as anger and disgust expressions induce more profound changes in the face appearance than the other expressions which make age prediction/perception more difficult. Our model clearly outperforms the existing methods [7, 9, 18] by a wide margin which further proves the effectiveness of our approach.

The hidden states capture the changes in the face appearance. To further illustrate this point, we show the face regions corresponding to each hidden state. More specifically, the averages of the bottom and top regions are computed (Figure 3). 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.

3.3 Joint-Learning for Expression Recognition

In this experiment, we consider a different, yet related, task: how age information can improve the recognition of expressions. Although aging affects how people exhibit expressions, much of automatic expression recognition methods do not use the age of the subject to recognize expressions. This is mainly due to the lack of expression datasets with a sufficiently large age range. Motivated by the introduction of recent age-expression datasets, Guo et al.

Figure 3: Average face regions corresponding to different hidden states (from left to right) for the bottom and top face regions.

Table 1: Expression-independent and expression-joint learning are evaluated on FACES and Lifespan datasets. The results show clear improvement of performance when the age is learnt jointly with the expression and the age prediction error is reduced by 14.43% and 37.75% for FACES and Lifespan datasets respectively. The results of the methods [7] and [13] along with the results using the joint learning method [9] are compared with ours. Our model obtains the best performance for both datasets with a large margin. Note that [7] and [13] assume that the expressions of the tested sample is a prior-knowledge while our model has no such requirement. The last column shows the difference in error when using the joint learning in comparison with independent learning.

Dataset	[7]	[13]	[9]	Indep-Learn	Joint-Learn	Reduc-Rate %
FACES	9.12	8.33	9.94	8.66	**7.41**	14.43%
Lifespan	6.63	6.23	8.85	8.45	**5.26**	37.75%

Table 2: Age estimation error for each expression subset on the FACES dataset. The error is reduced for all expressions using the expression-joint learning. The largest error reduction is achieved for neutral faces (30.13%) while the smallest error reduction is obtained for anger and disgust (4.64% and 2.43% respectively).

Test Data	Indep-Learn	Joint-Learn	Reduc-Rate %
Neutrality	8.54	**5.97**	30.13
Anger	8.61	**8.21**	4.64
Disgust	8.37	**8.17**	2.43
Fear	9.79	**8.25**	15.71
Happiness	8.42	**6.77**	19.58
Sadness	8.17	**7.07**	13.44
Average	8.66	**7.41**	14.43

[9] recently proposed a method to recognize facial expressions while reducing the influence of human aging.

We apply our model on FACES and Lifespan datasets to recognize the expression. The results are shown in Table 4. Our method improves the expression recognition performance for the FACES dataset by 2.38%. However, the accuracy on the Lifespan dataset is comparable to the one acquired by independent learning. This maybe explained by the observation that there are only two expressions in Lifespan compared to six ones in FACES, and hence the expression variation within Lifespan dataset is smaller than it is within the FACES dataset.

Table 3: Age estimation error for each expressions subset on the Lifespan dataset. The error is reduced for both neutrality and happiness expressions. Note that, since the numbers of happy and neutral faces are not equal, the weighted average is computed.

Test Data	Indep-Learn	Joint-Learn	Reduc-Rate %
Neutrality	8.66	**5.72**	33.94
Happiness	7.96	**4.14**	47.91
Average	8.45	**5.26**	37.75

Table 4: Expression recognition using age-joint and age-independent learning evaluated on FACES and Lifespan datasets. Joint-learning improves the accuracy by 2.38% on the FACES dataset while the accuracy on the Lifespan dataset is comparable. The method in [9] is further tested on our features, and the results show degrading in the performance for both datasets.

Dataset	Indep-Learn %	Joint-Learn %	[9] %
FACES	90.05	**92.19**	84.68
Lifespan	93.91	93.68	91.05

Consequently, the margin of improvement is smaller for the Lifespan dataset and the joint learning method obtains comparable accuracy.

We compare the proposed method with the one in [9]. As the authors manually labeled 31 fiducial points on the face and extracted the features using their locations, a direct comparison of the results will not be fair. Thus, we test the method in [9] using our features. The datasets are divided into the same four age groups ([18-29],[30-49],[50,69], and [70-94]). Then, a new class is created for each expression/age group combination resulting in 24 and 8 new classes for the FACES and Lifespan dataset respectively. The obtained accuracy (see Table 4) is lower than the one acquired by our model.

4 Discussion

The results obtained using our graphical model show the strength of joint-learning to alleviate the influence of facial expression in age prediction. Some existing works [7, 18] approached age prediction with variant facial expressions. Our method is different in two aspects: First, in our model, the age is jointly learnt with all expressions instead of learning the cross-expression for two expressions at a time. This property allows our model to be extended to a broader group of tasks where the changes are not restricted to the basic (profound) expressions. For example, the changes can be described by a group of smaller units (e.g. action units [5]). These changes can describe various face (undefined) expressions. In such cases, the hidden layer will learn the relationship between the age and multiple variables (action units) instead of one variable (expression) at a time. Moreover, beside facial expressions, other attributes can be learnt collectively within the proposed graphical model such as gender and race. Second, the proposed approach does not require the expression labels of the test samples to be known while the existing methods [7, 18] assume prior-knowledge of the expressions.

5 Conclusions

In this paper, an expression-invariant age predictor is proposed by jointly learning the age and the expression. We introduce a graphical model with a latent layer to learn the relationship between the age and the expression. This layer is designed to capture the changes in the face which induce the aging and the expression appearance.

Conducted on two age-expression datasets (FACES and Lifespan), our 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. Furthermore, using our model, without prior-knowledge of the expressions of the test faces, the acquired results are better that the best reported ones for both datasets.

Acknowledgement

This research is supported by the Dutch national program COMMIT. This research is partly supported by NICTA. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications, and the Digital Economy, and the Australian Research Council (ARC) through the ICT Centre of Excellence Program.

References

[1] The fg-net aging database, http://www.fgnet.rsunit.com.

[2] Sung Eun Choi, Youn Joo Lee, Sung Joo Lee, Kang Ryoung Park, and Jaihie Kim. Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognition*, 44(6):1262–1281, 2011.

[3] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169, 1985.

[4] 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.

[5] Paul Ekman and Wallace V. Friesen. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press*, 1978.

[6] Yun Fu, Guodong Guo, and Thomas S. Huang. Age synthesis and estimation via faces: A survey. *Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010.

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

[8] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S. Huang. Human age estimation using bio-inspired features. *Computer Vision and Pattern Recognition*, pages 112–119, 2009.

[9] Guodong Guo, Rui Guo, and Xin Li. Facial expression recognition influenced by human aging. *Affective Computing*, 4(3):291–298, 2013.

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

[11] Joris M. Mooij. Libdai: A free and open source c++ library for discrete approximate inference in graphical models. *Journal of Machine Learning Research*, pages 2169–2173, 2010.

[12] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[13] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. *Automatic Face and Gesture Recognition*, pages 341–345, 2006.

[14] 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.

[15] Zhiguang Yang and Haizhou Ai. Demographic classification with local binary patterns. *Advances in Biometrics*, pages 464–473, 2007.

[16] Chun-Nam John Yu and Thorsten Joachims. Learning structural svms with latent variables. *International Conference on Machine Learning*, pages 1169–1176, 2009.

[17] Alan L. Yuille and Anand Rangarajan. The concave-convex procedure (cccp). *Neural Computation*, pages 915–936, 2003.

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