A Comparative Evaluation of Regression Learning Algorithms for Facial Age Estimation

Carles Fernández[1], Ivan Huerta[2], and Andrea Prati[2]

[1] Herta Security
Pau Claris 165 4-B, 08037 Barcelona, Spain
carles.fernandez@hertasecurity.com
[2] DPDCE, University IUAV
Santa Croce 1957, 30135 Venice, Italy
huertacasado@iuav.it, aprati@iuav.it

Abstract. The problem of automatic age estimation from facial images poses a great number of challenges: uncontrollable environment, insufficient and incomplete training data, strong person-specificity, and high within-range variance, among others. These difficulties have made researchers of the field propose complex and strongly hand-crafted descriptors, which make it difficult to replicate and compare the validity of posterior classification and regression schemes. We present a practical evaluation of four machine learning regression techniques from some of the most representative families in age estimation: *kernel techniques*, *ensemble learning*, *neural networks*, and *projection algorithms*. Additionally, we propose the use of simple HOG descriptors for robust age estimation, which achieve comparable performance to the state-of-the-art, without requiring piecewise facial alignment through tens of landmarks, nor fine-tuned and specific modeling of facial aging, nor additional demographic annotations such as gender or ethnicity. By using HOG descriptors, we discuss the benefits and drawbacks among the four learning algorithms. The accuracy and generalization of each regression technique is evaluated through cross-validation and cross-database validation over two large databases, MORPH and FRGC.

Keywords: Age estimation, Support Vector Regression, SVM, Random Forest, Multilayer Neural Networks, Regularized Canonical Correlation Analysis, CCA, HOG

1 Introduction

Automatically conducting human age estimation from facial images can be valuable for a number of applications, including advanced video surveillance and biometrics [3, 8]; demographic statistics collection; business intelligence and customer profiling for targeted advertisements; and search optimization in large databases, to list some. Unfortunately, this problem has historically been one of the most challenging within the field of facial analysis. Some of the reasons are the uncontrollable nature of the aging process, the strong specificity to the personal traits of each individual [19], high variance of observations within the same age range, and the fact that it is very troublesome to gather complete and sufficient data to train accurate models [4].

The process of collecting quality age-annotated samples is difficult, and has often resulted in very limited and strongly skewed databases. This is especially disadvantageous for applications like video surveillance and forensics, which need to work correctly when facing unknown subjects and a lack of any additional cues. In these cases, the availability of large databases like MORPH [16] and FRGC [15] offers a great opportunity to make advances in the field. Keeping in mind that any training data set which is representative of the whole population cannot exist, the only viable option is to develop methods that are able to exploit large databases in order to gain substantial generalization capabilities. With these premises, this study includes the following contributions:

- A comparative evaluation of four of the most prominent machine learning regression techniques that have been typically applied to the problem of age estimation: *Support Vector Regression* (SVR), *Multilayer Neural Networks* (MNN), *Random Forests* (RF), and *Canonical Correlation Analysis* (CCA).
- The first attempt to use *Histograms of Oriented Gradients* [2] as a visual descriptor for age estimation. In our study, and compared to usual features, HOG benefits from being much faster to compute, standard, and easily replicable, besides offering very similar performance to the state-of-the-art.
- A baseline proposal for incorporating cross-database validation methodologies in order to test the generalization of an approach. To this end, we also propose the use of the large FRGC database, which has not received much attention in the past for age estimation purposes.

The paper is structured as follows: the next section reports and comments on previous works on age estimation. The proposed approach and methodology is described in Section 3. Section 4 analyzes current available age databases, describes the extensive experimentation carried out, and analyzes the obtained results. Finally, Section 5 draws some conclusions and gives hints for further research.

2 Related work

The first works and databases on automatic age estimation from digital images started appearing in the early 2000s [9–11]. Nonetheless, research in the field has experienced a renewed interest from 2006 on, since the availability of large databases like MORPH-Album 2 [16], which increased by 55$\times$ the amount of real age-annotated data with respect to traditional age databases. Next, we summarize the most successful descriptors and techniques that have been recently evaluated with this database.

Existing works on age estimation can be categorized by their choice of feature and classification scheme. Regarding visual features, shape and appearance models such as ASM (Active Shape Model) and AAM (Active Appearance Model) have been some of the primary cues used to model aging patterns [1, 4, 5, 9]. Such statistical models capture the main modes of variation in shape and intensity observed in a set of faces, and allow face signatures based on such characterizations to be encoded.

Bio-Inspired Features (BIF) [17] and its derivations have been consistently used for age estimation in the last years [4, 8]. These feed-forward models consist of a number

of layers intertwining convolutionally and pooling processes. First, an input image is mapped to a higher-dimensional space by convoluting it with a bank of multi-scale and multi-orientation Gabor filters. Later, a pooling step downscales the results with a non-linear reduction, typically a MAX or STD operation, progressively encoding the results into a vector signature. In [13], the authors carefully design a two-layer simplification of this model for age estimation by manually setting the number of bands and orientations for convolution and pooling. Such features are also used in their posterior works [6, 7].

With regards to the learning algorithm, Support Vector Machines (SVM) have commonly been used for age classification and regression, as in [13]. A binary decision tree with SVMs at each node is proposed in [8]. Age ranges are coarsely assigned, and later are more precisely estimated by Support Vector Regressors (SVR) at the leaves. In [1], a particular ranking formulation of support vectors, OHRank, is used. The approach uses cost-sensitive aggregation to estimate ordinal hyperplanes (OH) and ranks them according to the relative order of ages. In this paper AAM features are used. In [19], the author employs a similar ranking technique called MFOR.

There have been previous proposals training neural networks, which are able to learn complex mappings and deal with outliers, for age estimation. In [9], AAM-encoded face parameters are used as an input for the supervised training of a neural network with a hidden layer. In this case, models were trained uniquely from 200 color images, and the number of AAM-model parameters was restricted to 22. More recently, the authors of [4] tackle age estimation as a discrete classification problem using 70 classes, one for each age. The best algorithm proposed in this work (CPNN - Conditional Probability Neural Network) consists of a three-layered neural network, in which the input to the network includes both BIF features x and a numerical value for age y, and the output neuron is a single value of the conditional probability density function $p(y|x)$.

Although ensemble learning methods have not been extensively used in the field, they are particularly suitable for environments with high-dimensional features and strongly skewed data. An approach based on Random Forests (RF) over anthropometric measurements is presented in [12]. A collection of 11K simple features such as distances and area ratios are extracted from the facial mesh derived from 68 fine-annotated fiducial markers, and directly fed to the forest. The approach is tested using the subset of 710 face images from FG-NET that are between 0 and 20 years old.

Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA), along with their regularized and kernelized versions, are increasingly being used in the field of age estimation [6, 7]. These subspace learning algorithms were originally conceived to model the compatibility between two multidimensional variables. PLS uses latent variables to learn a new space in which such variables have maximum correlation, whereas CCA finds basis vectors such that the projections of the two variables using these vectors are maximally correlated to each other. Both techniques have been adapted for label regression. To the best of our knowledge, the best current result over MORPH is achieved by combining BIF features with kernel CCA [7], although in that case the size of training folds is limited to 10K samples due to computational limitations.

Our experiments demonstrate that a single-scale HOG descriptor is sufficiently expressive to compare with the performance of complex, fine-tuned features such as BIF; and that alignment through 5 fiducial points results in comparable age estimation per-

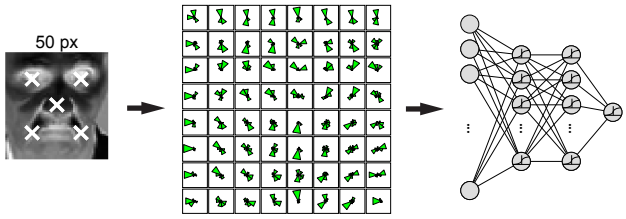


Fig. 1. The proposed study evaluates four machine learning regression techniques, all of them fed with histograms of gradients extracted from 50×50 pixel aligned images. A schematic example is shown here for the case of neural networks.

formance to precise alignment achieved through ASM/AAM fitting, which use approximately 70 landmarks. Additionally, our results show that the common strategy of partitioning age into ranges is unnecessary, since a single output neuron is able to act as an effective regressor. Posing age estimation purely as a regression problem not only simplifies the formulation, but additionally provides more accurate results when using a proper descriptor.

3 Methodology

In order to compare the performance of machine learning regression techniques in the field of age estimation, we have chosen a representative algorithm from each of four families that are conventionally used: a *Support Vector Regressor* from the kernel family, a *Multilayer Neural Network* from the neural network family, a *Random Forest* from the ensemble learning family, and a *regularized Canonical Correlation Analysis* from the projection family. These techniques are described next.

Preprocessing. In general, existing works tackle the problem of age estimation with visual features that are either complex and fine-tuned (e.g., BIF), or require precise statistical models involving tens of facial landmarks for accurate alignment (e.g., ASM or AAM models). As opposed to this, the four chosen learning algorithms will be evaluated with single-scale HOG visual descriptors after a 5-landmark alignment, see Figure 1. Histograms of Oriented Gradients (HOG) [2] have been largely used as robust visual descriptors in many computer vision applications related to object detection and recognition, due to their expressiveness, fast computation, compactness, and invariance to misalignment and monotonic illumination changes.

The facial region of each image has been extracted with the face detector described in [14]. Unlike previous approaches, we do not rely on precisely aligned appearance models. The relative alignment invariance of HOG allows us to require only five landmarks. The fiducial markers corresponding to the eye centers, nose tip and mouth corners have been obtained using the convolutional neural network for face alignment presented in [18]. The aligned version of each detected face is obtained by a non-reflective similarity image transformation that yields an optimal least-square correspondence be-

tween the set of fiducial points and the target locations, in which eye centers and mouth corners are symmetrically placed at 25% and 75% of the alignment template. Unlike previous approaches like [7], which use input images of 60×60 pixels, our aligned image are resized to only 50×50 pixels.

Support Vector Regression. Given a training set with L data examples $\mathbf{X} \in \mathbb{R}^m$ and their outputs $\mathbf{Y} \in \mathbb{R}^1$, the standard formulation of SVR under a given regularization cost C and slack variables ϵ is defined as

$$\begin{aligned} \min_{\mathbf{w},\mathbf{b},\xi,\xi^*} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\left(\sum_{i=1}^{L}\xi_i + \sum_{i=1}^{L}\xi_i^*\right) \\ s.t.\ \mathbf{w}^T\phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i \\ y_i - \mathbf{w}^T\phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i^* \end{aligned} \tag{1}$$

with $\xi_i, \xi_i^* \geq 0, i = 1 \ldots L$, where the kernel function $\phi(\mathbf{x_i})$ maps the feature vector $\mathbf{x_i}$ into a higher-dimensional space. This optimization problem is usually solved through its dual formulation, using algorithms such as Sequential Minimal Optimization.

Multilayer Neural Networks. The proposed MNN consists of an input visual feature layer, K hidden layers, and a single output neuron as an age regressor. All activation functions are set to be log-sigmoid, $\sigma(\beta, x) = \frac{1}{1+exp(-\beta x)}$. The prediction at layer k is

$$\mathbf{y}^{(k)} = \sigma(\mathbf{b}^{(k)} + \mathbf{\Theta}^{(k)}\mathbf{y}^{(k-1)}), \tag{2}$$

with $\mathbf{y}^{(1)} = \mathbf{b}^{(1)} + \mathbf{\Theta}^{(1)}\mathbf{x}$, where $\mathbf{x} \in \mathbf{X}$ is the input vector to the MNN, $\mathbf{b}^{(k)}$ a vector of unitary bias neurons and $\mathbf{\Theta}^{(k)}$ a matrix of connection weights. The cost function is

$$C(\mathbf{\Theta}) = \frac{1}{M}\left(-\mathbf{y}\log\mathbf{y}^{(K)} - (1-\mathbf{y}^{(K)})\log(1-\mathbf{y}^{(K)})\right) + \frac{\lambda}{2M}\sum_{k=1}^{K}{\mathbf{\Theta}^{(k)}}^2, \tag{3}$$

where $\mathbf{y} \in \mathbf{Y}$ is the ground truth and λ prevents overfitting by ℓ_2-regularization. Due to the log-sigmoid activations, the output regressor is factorized by 100 to directly provide age estimates in the range $[0 - 100]$. Each backpropagation step is accomplished by iterative unconstrained minimization of the multivariate cost function $C(\mathbf{\Theta})$, using

$$\frac{\partial\sigma}{\partial\theta_i} = \beta\sigma(\beta,\theta_i)\left(1-\sigma(\beta,\theta_i)\right) + \frac{\lambda}{M}\theta_i \tag{4}$$

as the multivariate gradient function for non-bias neurons. In our experiments, we used $\beta = 1$ and set the number of minimization iterations to 10.

Random Forests. Random forests are ensemble learners used for either classification or regression. The models are trained by applying bootstrap aggregation to N_{tree} base classification and regression trees. During the learning stage, each tree $h_n, n = 1, \ldots, N_{tree}$ samples with replacement a random selection of m-dimensional examples

$(\mathbf{X}_B, \mathbf{Y}_B) \in (\mathbf{X}, \mathbf{Y})$. For each new grown node, a number $m_{try} \ll m$ of predictor variables θ are randomly selected, and the variable that provides the best binary split over the bootstrapped subset according to some objective function is selected for that node. A tree is grown until leaf nodes are pure, i.e. they consist of samples containing a single label. When presented with new data, each tree individually predicts an output, and a collective target response is provided by averaging or majority voting of the forest $\{h_n(\mathbf{X}, \theta_n)\}, \forall n$, for regression and classification problems, respectively.

Regularized Canonical Correlation Analysis. CCA is posed as the problem of relating data $\mathbf{X}$ to labels $\mathbf{Y}$ by finding basis vectors w_x and w_y such that the projections of the two variables on their respective basis vectors maximize the correlation coefficient

$$\rho = \frac{{w_x}^T \mathbf{X}\mathbf{Y}^T w_y}{\sqrt{({w_x}^T \mathbf{X}\mathbf{X}^T w_x)({w_y}^T \mathbf{Y}\mathbf{Y}^T w_y)}}, \tag{5}$$

or, equivalently, finding $\max_{w_x, w_y} {w_x}^T \mathbf{X}\mathbf{Y}^T w_y$ subject to the scaling ${w_x}^T \mathbf{X}\mathbf{X}^T w_x$=1 and ${w_y}^T \mathbf{Y}\mathbf{Y}^T w_y$=1. For age estimation, labels in $\mathbf{Y}$ are unidimensional, so a least squares fitting suffices to relate these labels to the projected data features. Thus, only w_x is computed, by solving the following generalized eigenvalue problem:

$$\mathbf{X}\mathbf{Y}^T \left((1-\gamma_y)\mathbf{Y}\mathbf{Y}^T + \gamma_y I\right)^{-1} \mathbf{Y}\mathbf{X}^T w_x = \lambda \left((1-\gamma_x)\mathbf{X}\mathbf{X}^T + \gamma_x I\right) w_x \tag{6}$$

Regularization terms $\gamma_x, \gamma_y \in [0, 1]$ have been included in Eq. 6 to prevent overfitting. Although CCA admits extension to a kernelized version, in that case covariance matrices become computationally intractable with over 10K samples. In practice, regularized CCA works comparably to KCCA [7], it is much less computationally demanding, and will allow us to employ the same exact validation schemes over large databases.

4 Experimental Results

Age databases. Due to the nature of the age estimation problem, there is a restricted number of publicly available databases providing a substantial number of face images labeled with accurate age information. The most well known examples in the literature are the PAL database [11], with 580 frontal images from non-repeated subjects; the FG-NET Aging Database [10], with 1,002 face images from 82 subjects; the GROUPS database, with 28,231 faces of non-repeated subjects; the Face Recognition Grand Challenge v2.0 (FRGC) database [15], with 44,278 images from 568 subjects; and the MORPH II database [16], with 55,134 face images of 13,618 subjects.

PAL and FG-NET are comparatively negligible to the rest in terms of number of samples. Additionally, age annotations in GROUPS are discretized into seven age intervals, which makes it unsuitable for training accurate age regression models. Moreover, FG-NET contains only 82 subjects, so a *leave-one-person-out* validation scheme is employed by convention, to avoid optimistic biasing by identity replication. Given such limitations, and the recent tendency to use MORPH as a standard for age estimation,

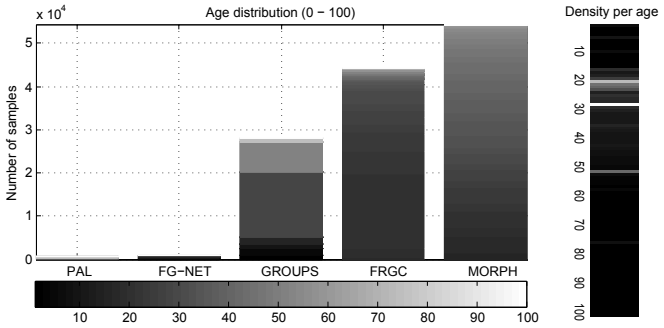


Fig. 2. Age distribution and density per database. In Age distribution the color represents the age. In Density per age the color represents the density (white color more density). PAL and FG-NET are relatively negligible compared to others, and GROUPS only provides age intervals, so we focus on MORPH II and FGRC. Age samples are mainly skewed towards people who are 20–30 and 50 years old.

we concentrate on this database and on FRGC to provide experimental evaluations. Although the FRGC database is comparable to MORPH regarding number of samples, image quality and age range coverage, we were not able to find any previous publication on age estimation including FRGC as part of their experiments. This new database for age estimation is described next. Figure 2 offers a graphical visualization and comparison of the analyzed databases, by number of samples and density of age ranges.

FRGC database [15] is presented in 2005 and contains approximately 50,000 images from 568 subjects. The database consists of four controlled still images, two uncontrolled still images, and one three-dimensional image for each subject session. Different sessions have been carried out during different years (2002, 2003, and 2004) with the same subjects. The controlled images were taken in a studio setting, are full frontal facial images taken under two lighting conditions (two or three studio lights). The uncontrolled images were taken in varying illumination conditions; e.g., hallways, atria, or outdoors. Each set of controlled and uncontrolled images contains two expressions, smiling and neutral. The 3D images were taken under controlled illumination conditions appropriate for the Vivid $900/910$ sensor. For our experiments we have used the database without the 3D images, only 44k images. They are 56% male, and 44% female, 69% White, 1% Black, and 30% Asian. The age range is between $18-70$ years old, with a proportion 56% between $18-22$, 21% between $23-27$, and 23% more than $28+$ years old.

Metrics. We adopt the conventional metrics of Mean Average Error (MAE) and Cumulative Score (CS) for comparison with recent literature. MAE computes the average age deviation error in absolute terms, $MAE = \sum_{i=1}^{M} |\hat{a}_i - a_i|/M$, with $\hat{a}_i$ the estimated age of the i-th sample (i.e. $y_i^{(K)}$ in the case of the MNN), a_i its real age and M the total number of samples. CS [1, 19, 8] is defined as the percentage of images for which the error e is no higher than a given number of years l, as $CS(l) = M_{e \leq l}/M$.

Related publications typically supply either an eleven-point curve for age deviations $[0-10]$, or simply the value $CS(5)$. We provide both results for future reference.

The optimal HOG parameters were searched for so as to minimize the MAE score over MORPH, using 5-fold cross-validation. In particular, the division into training and validation sets was made so that all the instances of the same subject were contained in one single fold at a time; this applies to all the experiments presented in this paper. We found that the version of 8×8 and 9 bins per histogram granted the best results. In our experiments, HOG descriptors are extracted directly from the aligned version of each detected face. We observed that the best results were achieved by directly inputting the per-cell, unitary-normalized HOG descriptors, without any further normalization. The selected HOG feature has the advantage of being considerably more compact than other features commonly used in the literature. For instance, the BIF feature in [13] is 4376-dimensional, whereas ours is only 576. This results in faster convergence of the algorithm, less data complexity (deriving into fewer layers required), and a smoother regularization parameter space for a similar expressive power.

SVR experiments were carried out using the ϵ-SVR implementation from the LIBSVM library[3], with a Radial Basis Function kernel. All of our experiments employ the same input features, i.e. the signed version of unit-normalized HOG over an 8×8 grid and 9 encoding bins. The optimal regularization cost C, and hyper-parameters γ and ϵ have been independently obtained for each target database, through exhaustive logarithmic grid search and 5-fold cross-validation. A similar grid search has been carried out for RF, in order to adjust the optimal parameters N_{tree} and m_{try}. In this case, the technique was particularly invariant to the choice of parameters.

The optimal architecture and regularization for the MNN have been explored through grid-search again, dividing the target database into 20 folds. Each MORPH subset contained over 3K samples, granting faster convergence.

Regarding CCA, only the regularization terms need to be optimized. Interestingly, the best choice of regularization turned out to be $\gamma_x = \gamma_y = 0$. This is explained by the size of the descriptor, which is orders of magnitude smaller than the number of examples ($576 \ll L$), and thus less prone to overfit the data.

Table 1 shows a thorough comparison with publications that supply cross-validation MAE using MORPH (referred as MORPH-5CV). Unlike many previous works [1, 6, 7, 19], our MORPH-5CV experiment exploits the whole available set of 55K samples, by training from 4 folds, testing over the remaining one and averaging all five combinations.

Table 2 includes scores for FRGC as a future reference for the research community (FRGC-5CV). Furthermore, we feel that cross-database (CDB) generalization would be a challenging and interesting metric to take into account for validating the robustness of a method. For this reason, we additionally include the generalization scores obtained when training a whole database using its optimal 5CV parameters, and testing it completely with the other database. These appear in the aforementioned table as MORPH→FRGC and FGRC→MORPH, where the testing database appears last. These results also prove the good generalization properties of the evaluated techniques.

[3] `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

	MORPH-5CV			
Technique	Feature	Train/test	MAE	CS(5)
WAS [10, 4]	AAM+BIF	55K	9.21	–
AAS [4]	AAS+BIF	55K	10.10	–
AGES [5, 4]	AAM+BIF	55K	6.61	–
RED (SVM) [1]	AAM	6K	6.49	48.9%
OHRank [1]	AAM	6K	6.07	56.4%
OHRank [1, 4]	AAM+BIF	55K	6.28	–
PLS [6, 7]	BIF	10K/55K	4.56	–
kPLS [6, 7]	BIF	10K/55K	4.04	–
IIS-LLD [4]	AAM+BIF	55K	5.67	–
CPNN [4]	AAM+BIF	55K	4.87	–
CCA [7]	BIF	10K/55K	5.37	–
rCCA [7]	BIF	10K/55K	4.42	–
kCCA [7]	BIF	10K/55K	3.98	–
MFOR [19]	PCA+LBP+BIF	4K	4.20	72.0%
SVM+SVR [8]	BIF+ASM	78K	4.20	72.4%
SVR	**HOG**	55K	4.83	63.4%
MNN	**HOG**	55K	7.91	34.0%
RF	**HOG**	55K	6.84	43.1%
rCCA	**HOG**	55K	4.84	64.1%

Table 1. Age estimation results in MORPH II for the compared algorithms and visual descriptors, in a variety of settings. Symbol (–) indicates unreported values.

	MAE				CS(5)			
Validation scheme	SVR	ANN	RF	CCA	SVR	ANN	RF	CCA
MORPH–5CV	**4.83**	7.91	6.84	**4.84**	**63.4%**	34.0%	43.1%	**64.1%**
FRGC–5CV	**2.88**	4.34	4.08	4.41	**89.2%**	71.0%	82.1%	73.7%
MORPH→FRGC	7.55	7.47	8.92	**7.43**	41.3%	43.1%	30.5%	**44.8%**
FRGC→MORPH	**8.89**	**8.90**	9.51	8.50	**39.1%**	36.0%	33.6%	**39.1%**

Table 2. MAE scores and CS(5) percentages from all four classification schemes for MORPH and FRGC, under the two validation scenarios.

Figure 3 shows CS curves for the MORPH database, comparing our four evaluations with different published algorithms: BT, BP, kNN, SVM, SVR and RED-SVM [1]; OHRank[1]; MFOR [19]; and SVM+SVR [8]. In case of algorithm variations, best curves were chosen. Likewise, Figure 4 shows the CS curves of our proposed algorithms for MORPH and FRGC, both for 5CV and for the more challenging CDB validation scenario. Table 2 details CS(5) scores for these validation scenarios. For those works that facilitated the CS(5) score for the MORPH database, this metric has also been included for comparison in Table 1.

When comparing the accuracy of the techniques, we see that CCA and SVR perform similarly well, and comparable to state-of-the-art algorithms employing much more complex features and preprocessing. We have observed that MNN is very sensitive to

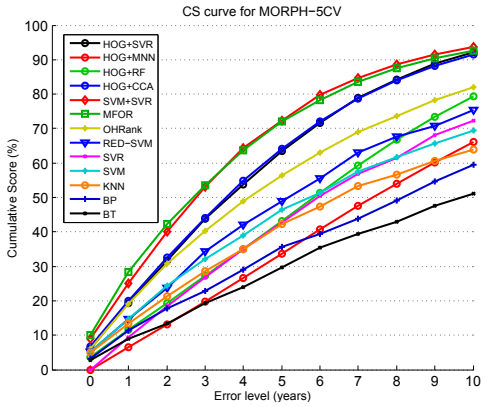
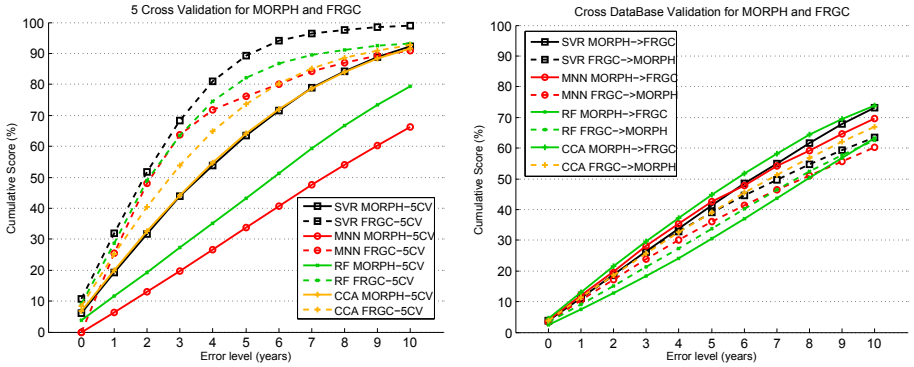


Fig. 3. Comparative of Cumulative Score curves from recent age estimation algorithms and our four evaluations: HOG+SVR, HOG+MNN, HOG+RF, HOG+CCA.



(a) (b)

Fig. 4. Cumulative Score curves of the four evaluated techniques, for MORPH and FRGC, in (a) 5-fold cross-validation (5CV) and (b) cross-database validation (CDB) scenarios.

particular weight initializations, RF is quite invariant to its parameterization. Regarding computational efficiency during training, the time lapse for a cross-validation fold is approximately 9 hours for SVR, 5 hours for RF, 9 min for MNN and 3 sec for CCA, on an Intel i7 computer at 1.6 GHz. This gives significant advantage to CCA.

5 Conclusions

We evaluated four machine learning techniques applicable to age regression from facial images: SVR, MNN, RF and CCA. We demonstrated that replacing complex feature extraction schemes with HOG features achieves comparable performance to the state-of-the-art, while being faster and easily replicable. Our approach requires less feature

tuning; it does not involve statistical face models requiring precise annotation of tens of facial landmarks; and it does not require additional cues. During the method comparison, CCA and SVR similarly provided the best accurate results, although the combination of HOG+CCA proved to be the most computationally efficient and straightforward, not even requiring parameter adjustment. Furthermore, we introduced FRGC as a suitable (and so far unnoticed) large database for age estimation, and proposed a cross-database validation scheme to test the generalization of age estimation methods. Further research should explore the incorporation of additional cues such as gender and ethnicity. These cues have been effectively used in the past to increase age estimation accuracy [7].

Acknowledgments. The authors would like to thank to Christina Zitello for English editing. This work has been partially supported by the Spanish Ministry of Science and Innovation (MICINN) through the Torres-Quevedo funding program (PTQ-11-04401).

References

1. Kuang-Yu Chang, Chu-Song Chen, and Yi-Ping Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *CVPR*, pages 585–592. IEEE, 2011.
2. Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893. IEEE, 2005.
3. Yun Fu, Guodong Guo, and T.S. Huang. Age synthesis and estimation via faces: A survey. *TPAMI*, 32(11):1955–1976, 2010.
4. Xin Geng, Chao Yin, and Zhi-Hua Zhou. Facial age estimation by learning from label distributions. In *TPAMI*, volume 35, pages 2401–2412. IEEE, 2013.
5. Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age estimation based on facial aging patterns. *TPAMI*, 29(12):2234–2240, 2007.
6. Guodong Guo and Guowang Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *CVPR*, pages 657–664. IEEE, 2011.
7. Guodong Guo and Guowang Mu. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In *10th Int. Conf. on Automatic Face and Gesture Recognition*. IEEE, 2013.
8. Hu Han, Charles Otto, and Anil K Jain. Age estimation from face images: Human vs. machine performance. In *International Conference on Biometrics (ICB)*. IEEE, 2013.
9. Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. Comparing different classifiers for automatic age estimation. *TSMC-B*, 34(1):621–628, 2004.
10. Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *TPAMI*, 24(4):442–455, 2002.
11. Meredith Minear and Denise C Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.
12. Albert Montillo and Haibin Ling. Age regression from faces using random forests. In *ICIP*, pages 2465–2468. IEEE, 2009.
13. Guowang Mu, Guodong Guo, Yun Fu, and Thomas S Huang. Human age estimation using bio-inspired features. In *CVPR*, pages 112–119. IEEE, 2009.
14. David Oro, Carles Fernández, Javier Rodríguez Saeta, Xavier Martorell, and Javier Hernando. Real-time GPU-based face detection in HD video sequences. In *ICCV Workshops*, pages 530–537, 2011.

15. P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, pages 947–954. IEEE, 2005.
16. K. Ricanek and T. Tesafaye. MORPH: a longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition*, pages 341–345, 2006.
17. Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.
18. Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, pages 3476–3483. IEEE, 2013.
19. Renliang Weng, Jiwen Lu, Gao Yang, and Yap-Peng Tan. Multi-feature ordinal ranking for facial age estimation. In *AFGR*. IEEE, 2013.