

Age Recognition in the Wild

Amirhossein Jahanbekam^{1,3},
¹RWTH-Aachen University
 52056 Aachen, Germany
<http://www.rwth-aachen.de>

Christian Bauckhage^{2,3}
²B-IT, University of Bonn
 53113 Bonn, Germany
<http://www.b-it-center.de>

Christian Thureau³
³Fraunhofer IAIS
 53754 St. Augustin, Germany
<http://www.iais.fraunhofer.de>

Abstract

In this paper, we present a novel approach to age recognition from facial images. The method we propose, combines several established features in order to characterize facial characteristics and aging patterns. Since we explicitly consider age recognition in the wild, i.e. vast amounts of unconstrained Internet images, the methods we employ are tailored towards speed and efficiency. For evaluation, we test different classifiers on common benchmark data and a new data set of unconstrained images harvested from the Internet. Extensive experimental evaluation shows state of the art performance on the benchmarks, very high accuracy for the novel data set, and superior runtime performance; to our knowledge, this is the first time that automatic age recognition is carried out on a large Internet data set.

1. Motivation and background

Human age recognition is recently getting more attention due to its practical relevance in many real-world applications, including security, biometrics, as well as entertainment technology. In this paper, we present a novel efficient approach to age recognition, that considers classification of a wide range of patch-based age sensitive features.

In the literature, aging effects on human images have been studied within 3 different contexts: **(i) Age simulation:** [12] studies the cranofacial growth of young faces and models the anthropometric features of faces for a facial ontogeny. **(ii) Age transformation:** using Gabor filters or shape and texture models, the work in [13, 11] tries to alleviate aging effects in order to improve face recognition. **(iii) Age estimation:** most related work address the problem of age estimation. [18] employs Local Binary Pattern Histograms (LPBH) to identify 4 age groups. [16] uses patch distributions to encode visual features of a face. [14] considers a hier-

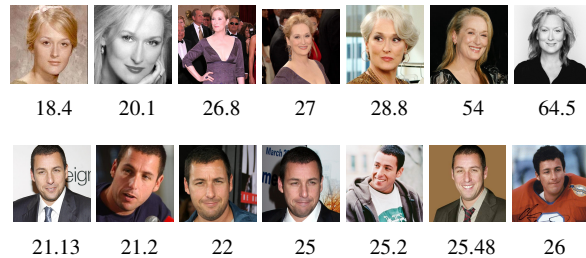


Figure 1. Images ranked according to age.

archy of 3 levels of facial features with four embedded feature types and [6] applies a pyramid of Gabor filters to obtain a wrinkle descriptor.

Note that most previous work focuses on *exact age estimation*. However, exact age estimation is an ill-posed problem. Consider, for example, the celebrity pictures shown in this paper. Typically, people strive to look younger than they really are. Moreover, human observers commonly over- or underestimate other people's age. Often, we can only reliably tell if a person looks older or younger than somebody else. Therefore, this paper addresses *relative age estimation*. Typical results of exact and relative/ranked age estimates obtained from our approach are shown in Fig. 1.

2. Features

Figure 2 provides an overview of our approach. Given an image, we use a standard face detector to extract a facial image, rescale it to uniform size, and convert it to gray values. According the successful strategy of [14, 16, 17, 20], we, too, base our approach on the idea of bags of local descriptors. To this end, we superimpose two overlapping, regular grids over a face image and compute a suit of local features for each of the resulting smaller patches. We found that by combining different feature descriptors, we could account for more

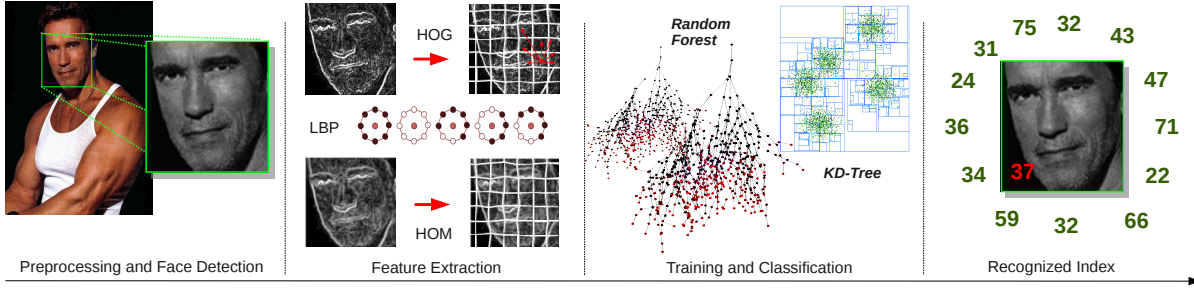


Figure 2. Overview of the age estimation approach

age-relevant information of a facial image. In detail, we combined four feature-descriptors:

(i) Local gradient statistics [1, 4] were reported to be particularly successful in age recognition [6]. For describing local gradient statistics, we use a histograms of oriented gradients (HOGs) [4] descriptor (with 10 bins). However, we found that the raw gradient magnitude (as opposed to normalized gradient magnitudes in the HOG descriptor) appears to characterize local facial texture and wrinkles. This is perfectly in line with the work in [8] which revealed that gradient magnitude should be discarded for age independent face recognition. Thus, since we are interested in age dependent classification, the gradient magnitude is used as an additional cue.

(ii) We found that extreme local intensity differences carry valuable information for a person’s age, as they often appear to be an indicator for wrinkles. Thus, we consider a min/max range filter

$$g_{ij} = \max(f_{kl}) - \min(f_{kl})$$

where $k \in \{i-1, i, i+1\}$ and $l \in \{j-1, j, j+1\}$. This filter is an edge detector that emphasizes intensity differences in the neighborhood of a pixel. Similar to the HOG descriptor, we compute histograms of range filter responses for each of the patches (using 5 histogram bins). Here, however, we consider only their magnitudes and discard directional information.

(iii) We include *Local Binary Patterns* (LBP) for providing a computationally cheap texture descriptor (following ideas suggested in [18]). Commonly, LBPs are aggregated into a histogram of 256 bins. In contrast, we found that a lower number of histogram bins (usually only 8) lead to a better generalization while still maintaining sufficient discriminative power.

(iv) Similar in spirit to the spatially flexible patches introduced in [16, 17], we further extend our patch descriptors by location information. In addition to the 10 dimensional histograms of gradient orientation, 5 dimensional histograms of gradient magnitudes and the 8 dimensional of LBP, we include 2 more dimensions

into the descriptor which characterize the location of the corresponding patch with respect to the center point of the face image.

3. Classification

We assume that a set $\{\vec{x}_i\}$ of $i = 1, \dots, n$ feature vectors is extracted from a face image. The probability of that person having a certain age a conditioned on these observations is given by

$$\begin{aligned} p(a|\vec{x}_1, \dots, \vec{x}_n) &= \frac{p(\vec{x}_1, \dots, \vec{x}_n|a) p(a)}{p(\vec{x}_1, \dots, \vec{x}_n)} \\ &= \alpha p(\vec{x}_1, \dots, \vec{x}_n|a). \end{aligned} \quad (1)$$

It is known, that images of faces and descriptors of facial features form manifolds in their respective features space [2, 19]. In terms of a probabilistic model, geometric constraints like this can be accounted for through latent variables. We therefore assume the following model

$$\begin{aligned} \alpha p(\vec{x}_1, \dots, \vec{x}_n|a) &= \alpha \sum_j p(\vec{x}_1, \dots, \vec{x}_n|c_j) p(c_j|a) \\ &= \alpha \sum_j \prod_i p(\vec{x}_i|c_j) p(c_j|a) \\ &= \sum_j w_j p(c_j|a) \end{aligned} \quad (2)$$

which divides the feature space into cells c_j . Each feature vector has a probability of residing in a cell and the cells are conditioned on the age variable.

Inference with this model is rather simple, because the term $w_j = \alpha \prod_i p(\vec{x}_i|c_j)$ in (2) can be estimated from voting over a quantized parameter space [15]. In our implementation, we therefore compute an appropriate quantization of the feature space and estimate the quantities $p(c_j|a)$ using a large set of training samples. Then, in the recognition phase, each member of a set of newly extracted feature vectors \vec{x}_i votes for the cells

Table 1. Features summary of FG-NET, UT-D and Internet databases.

Database	No. Images	Age Range	Age distribution	Image quality	Color
UT-Dallas	580	18 - 93	Not monotonic	Rather high	Colorful
FG-Net	1000	0 - 68	Not monotonic	Mid & Low	Monochrome
Internet Database	50,000	Not specified	Not Specified	High, Mid & Low	Colorful & Monochrome

c_j . This yields weights w_j which can be plugged into (2) to obtain $p(a|\vec{x}_1, \dots, \vec{x}_n)$. The corresponding age estimate for face image is then the age a , for which this expression assumes a maximum.

In order to obtain a data specific quantization of the feature space, we apply k D-trees or random forests. These allow for highly efficient retrieval and nearest neighbor searches [3, 7, 9]. In the training phase of our system, we grow a k D-tree/random forest given a large set of labeled training samples. The resulting subdivision of the features space defines the cells c_j which we consider in our latent model. Since the training samples are labeled, we can also estimate the densities $p(c_j|a)$ at this stage.

In the application phase, the tree structure is used to rapidly assign an input vector \vec{x}_i to a cell c_j and thus to vote for that cell. After all vectors \vec{x}_i extracted from an input image have been processed in this manner, age classification proceeds as stated above.

4. Experiments and evaluation

We trained our system using images from the UT-Dallas database [10] and the FG-Net database [5]. For testing, we also considered an additional, large database of portraits of celebrities (mostly of Caucasian and African origin) harvested from the Internet. Recognition accuracies we obtained from tests with the FG-Net and UT-Dallas data are similar and match the performance of state of the art approaches. However, because of known weaknesses in the FG-Net database [14], here, we only discuss our the results for the UT-Dallas data set. See Table 1 for more details on the data sets.

(i): experimenting with the UT-Dallas and the FG-Net data set, we randomly selected 250 images for training and used the remaining data for testing. Varying the patch size considerably influenced recognition accuracy and processing speed, as can be seen in Fig. 3 and Table 2. We found that a size ratio of patch to face image of 1/6 or 1/5 provides sufficient recognition accuracy at

Table 2. Pre-Processing time with regard to face size and patch size ratio.

Face Size	60	60	60	60	120	120	120	120	180	180	180
Patch Size	6	10	12	15	12	20	24	30	30	36	45
Ratio	1/10	1/6	1/5	1/4	1/10	1/6	1/5	1/4	1/6	1/5	1/4
# Patches	381	121	81	49	381	121	81	49	121	81	49
Time (ms)	1260	395	270	185	1580	620	440	380	870	730	610

almost real-time performance. For feature extraction, most computing time is spent on the computation of the magnitude of the HOG descriptors. For time critical applications, skipping this step leads to an accuracy decrease of 2% but considerably lowers computational costs. In our tests, we found that computation times for standard desktop computers can be reduced by 100ms on average.

Figure 3(b) shows the effects of varying the parameters in computing k d-trees or random-forests. The experiments indicate that for efficient processing and accurate results, the number of trees should vary between 30 to 100. Moreover, we were interested in the importance of the different features with respect to recognition accuracy. It appears that the vertical components of the local gradient features (HOG descriptors) are the most informative cue. The remaining feature descriptors show a (more or less) equally distributed relevance.

Figure 3(c) shows results that compare the performance of k d-tree and random forests. It can be seen that the resulting accuracy is almost identical. However, random forests are to prefer as they are usually faster to evaluate. In our implementation, a single run takes about 200 ms; an SVM, on the other hand, requires up to 90s. The dense average of our mean absolute errors on the FG-Net and UT-Dallas data sets are 7.54 and 6.47, respectively. Therefore, the prediction accuracy we achieve with our approach is comparable to that of other recent approaches but its computation times are usually much faster.

(ii) Our experiments using the Internet celebrity data were evaluated in a different manner. From a set of 50.000 Internet images (which do not contain any age related annotations), we randomly selected 100 pairs of images where each pair contained two pictures of the same person. Then, we applied the proposed approach to the images and let it sort them according to the estimated age. For verifying the results, we asked human subjects to verify if these *relative age estimation* results were correct; i.e. the task was to verify whether the

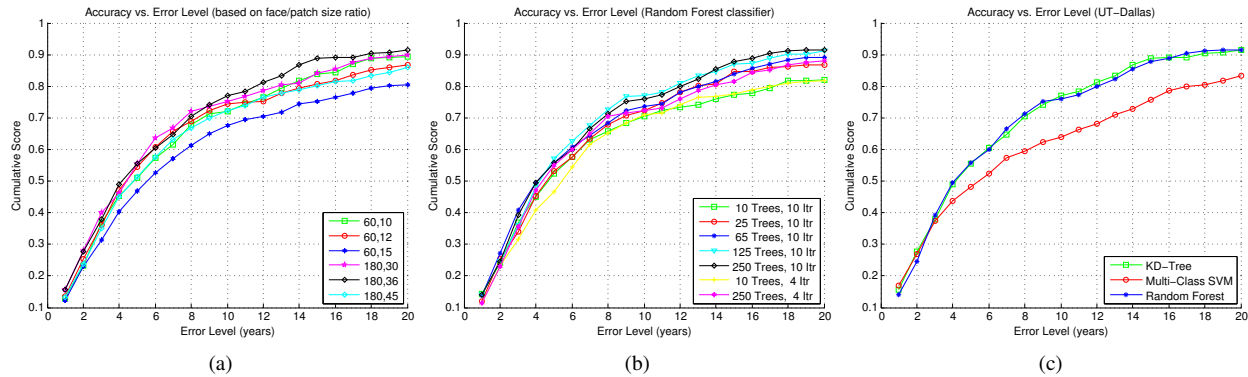


Figure 3. Accuracy of different classifiers

algorithm correctly ordered the two images. We measured an accuracy of 77%. However, a closer inspection revealed that cases of misclassification were mainly due to distorted perspective (people looking sideways etc.). Therefore, in a second experiment, we restricted the candidate images to strictly frontal views of faces. We found that this improved the classification accuracy to 85%. Or, stated in more informal terms, given two images of the same person, our approach is able to discern which of the two snapshots shows the younger/older face (according to human standards) in 85% of all cases.

5 Conclusion

This paper presented an efficient and accurate approach to human age recognition from facial images. We combined various recent feature descriptors and evaluated their use for automatic age recognition. Evaluation on established public benchmark data, as well as on a novel large and unconstrained data set showed a high accuracy combined with real-time performance. This makes the approach especially suitable for the end-consumer market using hand-held devices, e.g digital-cameras or cell phones.

References

- [1] M. Abdel-Mottaleb. Image Retrieval Based on Edge Representation. In *ICIP*, 2000.
- [2] O. Arandjelovic, G. Shakhnarovic, J. Fisher, R. Cipolla, and T. Darrell. Face Recognition with Image Sets Using Manifold Density Divergence. In *CVPR*, 2005.
- [3] O. Boiman, E. Shechtman, and M. Irani. In Defence of Nearest-Neighbor Image Classification. In *CVPR*, 2008.
- [4] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR*, 2005.
- [5] Face and Gesture Recognition Research Network. FG Net Aging Database. <http://www.fgnet.rsunit.com>.
- [6] G. Guo, G. Mu, Y. Fu, and T. Huang. Human Age Estimation using Bio-inspired Features. In *CVPR*, 2009.
- [7] J. Kubica, J. Masiero, A. Moore, R. Jedicke, and A. Connolly. Variable KD-Tree Algorithms for Spatial Pattern Search and Discovery. In *NIPS*, 2005.
- [8] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs. A study of face recognition as people age. In *ICCV*, 2007.
- [9] K. Mikolajczyk and J. Matas. Improving Descriptors for Fast Tree Matching by Optimal Linear Projection. In *ICCV*, 2007.
- [10] M. Minear and D. Park. A Lifespan Database of Adult Facial Stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.
- [11] U. Park, Y. Tong, and A. Jain. Age-invariant Face Recognition. *IEEE Trans. PAMI*, 32(5):947–954, 2010.
- [12] M. Ramanathan and R. Chellappa. Modeling Age Processing in Young Faces. In *CVPR*, 2006.
- [13] R. Singh, M. Vesta, A.Noore, and S.Singh. Age transformation for improving face recognition performance. *PREMI*, 4815 of LNCS:576–583, 2007.
- [14] J. Suo, T. Wu, S. Zhu, and W. G. S. Chen, X. Chen. Design Sparse Features for Age Estimation Using Hierarchical Face Model. In *FG*, 2008.
- [15] N. Toronto, B. Morse, D. Ventura, and K. Seppi. The hough transform’s implicit bayesian foundation. In *ICIP*, 2007.
- [16] S. Yan, M. Liu, and T. Huang. Extracting Age Information from Local Spatially Flexible Patches. In *ICASSP*, 2008.
- [17] S. Yan, H. Wang, J. Liu, X. Tang, and T. Huang. Ranking with Uncertain Labels and its Applications. In *ICME*, 2007.
- [18] Z. Yang and H. Ai. Demographic Classification with Local Binary Patterns. In *Int. Conf. Adv. in Biometrics*, 2007.
- [19] F. Yun, Y. Xu, and T. Huang. Estimating Human Age by Manifold Analysis of Face Pictures and Regression on Aging Features. In *ICME*, 2007.
- [20] X. Zhuang, X. Zhou, M. Hasegawa-Johnson, and T. Huang. Face Age Estimation Using Patch-based Hidden Markov Model Supervector. In *ICPR*, 2008.