Image and Vision Computing 30 (2012) 946–953

Contents lists available at SciVerse ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions[☆]

Fares Alnajar [a,b,*], Caifeng Shan [b], Theo Gevers [a], Jan-Mark Geusebroek [a]

[a] Informatics Institute, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
[b] Philips Research, High-Tech Campus 36, Eindhoven 5656AE, The Netherlands

ARTICLE INFO

Article history:
Received 2 November 2011
Received in revised form 5 April 2012
Accepted 30 July 2012

Keywords:
Age estimation
Codebook
Local features
Face

ABSTRACT

In this paper we propose to adopt a learning-based encoding method for age estimation under unconstrained imaging conditions. A similar approach [Cao et al., 2010] is applied to face recognition in real-life face images. However, the feature vectors are encoded in hard manner i.e. each feature vector is assigned to one code. The face is divided into patches where a code histogram is built for each patch. However, the codebook is learned using sample features from the entire face.

Therefore, we propose an approach to extract robust and discriminative facial features and use soft encoding. Instead of learning a codebook from the entire face, we extract and learn multiple codebooks for individual face patches. The encoding is done by a weighting scheme in which each pixel is softly assigned to multiple candidate codes. Finally, orientation histogram of local gradients in neighborhood has been introduced as feature vector for code learning.

On a large scale face dataset which contains 2744 real-life faces, the age group classification using our method achieves an absolute(relative) improvement of 3.6%(6.5%) over the best reported results [Shan, 2010].

1. Introduction

Automatic age estimation of a person is an interesting and challenging task, with many important applications in human–computer interaction, market intelligence and visual surveillance. Since human faces provide most information to perceive the age, most previous research efforts have focused on age estimation from face images [1].

Constructing a proper face image representation is a key component for successful face age estimation systems. Typically two kinds of features are extracted from face images: appearance features (e.g. wrinkles, skin roughness) and geometric features (e.g. shapes, ratios of distances between facial landmarks). For applications where images are acquired in unconstrained settings, it is difficult to automatically detect a sufficient number of fiducial landmarks to compute the geometrical features of the face.

As reviewed in [1], many approaches have been exploited to represent and model faces from images such as anthropometric models, age subspace or manifold, and active appearance models. However, each representation has its limitations and strengths. For example, the anthropometric model is useful for young ages, but not appropriate for adults; for age manifold learning, a large number of training samples is needed. The facial representation should not only be discriminative but also robust to appearance variations and noise. In recent years, local descriptor based approaches have been proven to be effective for face image analysis [4–7]. Traditionally, Gabor-wavelets have widely been exploited to model local facial appearance [4,8]. Recently, the histogram of Local Binary Patterns [5] has been adopted to describe the micro-structures of the face [9–11]. Tolerance against monotonic illumination changes and computational simplicity are the most important properties of LBP features. Scale-Invariant Feature Transform (SIFT) [6] and Histogram of Oriented Gradients (HOG) [7] are other types of local descriptors that have shown good performance in face analysis [12] and object recognition.

More recently, Cao et al. [13] argued that these local descriptors use manually designed encodings, and it is difficult to get an optimal encoding method. As shown in [13], the existing handcrafted codes are unevenly distributed, and some codes may rarely appear in face images. This means that the resulting code histogram is less informative and less compact. They used a learning-based encoding method, which adopts unsupervised learning methods to encode the local micro-structures of the face into a set of discrete codes. With Principal Component Analysis (PCA) and normalization, their learning-based descriptor achieves superior performance on face verification. Instead of face verification, in this paper, we consider learning-based encoding in the context of age estimation.

We adopt the learning-based encoding method for age estimation and propose an approach of extracting robust and discriminative facial features and encoding. First, instead of learning a codebook

[☆] This paper has been recommended for acceptance by Ming-Hsuan Yang.

[*] Corresponding author at: Informatics Institute, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands. Tel.: +31 20 525 7465; fax: +31 20 525 7490.

E-mail address: F.Alnajar@uva.nl (F. Alnajar).

0262-8856/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.imavis.2012.07.009

from the entire face, we extract and learn multiple codebooks for individual face patches. The intuition behind this is that the features histogram is computed for each patch. Second, the encoding is done by a weighting scheme in which each pixel is softly assigned to multiple candidate codes. This is to alleviate ambiguity especially in noisy real-life images. Aging effects are mainly observed as textural variations in faces such as wrinkles and other skin artifacts. Therefore, we investigate the use of orientation histogram of local gradients to describe faces for age estimation.

The rest of the paper is organized as the following. In Section 2 we provide an overview on related work. Section 3 describes learning-based encoding method. We outline our adaptations in Section 4. Experiments are presented in Section 5. Section 6 concludes the paper.

2. Related work

In the last few years, many research efforts have been invested on age estimation from face images. A thorough survey of the state of the art can be found in [1].

Geng et al. [14] introduce the Aging Pattern Subspace for age estimation, where an aging pattern is defined as a sequence of face images from the same person, sorted in the temporal order. This approach is evaluated on the FG-NET aging database, achieving a Mean Absolute Error (MAE) of 6.77 years. However, in general, it is difficult to collect multiple face images of the same person at different ages. Instead of learning a specific aging pattern for each individual, a common aging pattern could be learned from face images of multiple people [15]. Manifold learning techniques are adopted to embed face images into a low-dimensional aging manifold. The age manifold based regression [16] produces a MAE of 5.07 years on the FG-NET aging database.

Further, Yan et al. [17,18] propose to use Spatially Flexible Patches as face representation. This technique considers local patches and information about the position. Modeled by a Gaussian mixture model, their approach achieves a MAE of 4.95 years on the FG-NET database. Guo et al. [20] introduces the Biologically Inspired Features for age estimation. Combined with SVM, the proposed features produce a MAE of 4.77 years on the FG-NET database. Recently Ni et al. [21] collected a large web image database, and built a universal age estimator based on multi-instance regression.

Yang and Ai [10] consider LBP features for age estimation. They achieve the error rate of 7.88% on the FERET database and 12.5% on the PIE database. Further, Gao and Ai [8] study the problem of age estimation in consumer images. In their approach, Gabor features are extracted and used with Linear Discriminant Analysis (LDA). They consider four age categories: baby (0–1), child (2–16), adult (17–50), and old (50+). Trained on 5408 faces, their age estimator achieved an accuracy of 91% on 978 testing images. Gabor features are demonstrated to be more effective than LBP features and pixel intensities in their study. More recently, Shan [22] applies Adaboost to learn local features, both LBP and Gabor features, for age estimation on real-life faces acquired in unconstrained conditions.

Cao et al. [13] use a learning-based encoding method, which adopts unsupervised learning methods to encode the local microstructures of the face into a set of discrete codes. The method achieves high accuracy for face verification. It should be mentioned that other previous works proposed learning local descriptor. Meng et al. [27] used Local Visual Primitives (LVP) for face modeling and recognition. LVP is a representative of a face patch which appears frequently. Xie et al. [28] proposed to learn local Gabor Patterns for face representation and recognition.

In the next section, we extend this method to age estimation. The extension consists of three points: the codes are assigned in soft manner, different codebooks for different face patches, and using features more related to estimating the age.

3. Learning-based encoding

In this section, we briefly describe the learning-based encoding method [13]. At each pixel, its neighboring pixel intensities are sampled in a ring-based pattern to form a low-level feature vector. $r*8$ values are sampled at even intervals on the ring of radius r. The authors extensively varied the parameters (e.g. ring number, ring radius, sampling number of each ring), and found that the differences among patterns are not of influence on the face database they used. Following [13], we use the second sampling method with two rings ($r=1$, $r=2$, with center), that is, 25 values (8 from the first ring, 16 from the second ring, and the center value). After sampling, the sampled feature vector is normalized into unit length, to make the feature vector invariant to local illumination changes.

Then, the encoder is learned by applying unsupervised learning to a set of training face images. The feature vectors are extracted at each pixel. Different unsupervised learning methods are considered. In [13], three methods are examined: K-means, PCA tree, and random-projection tree [23]. Their experiments show that the difference among these learning schemes is small. In this paper, the PCA tree [23] is adopted. The largest principal component for the vectors at each node is first computed. After projecting the vectors onto that principal component, the vectors are split from the median value and two children nodes are created; the principal component and the median value are stored in the parent node. These children nodes are further split until the leaf number is equal to the code number, where each leaf represents one code. With the learned encoder, the input face image is encoded. Similar to LBP features, the encoded face image is divided into a grid of patches (7×5 patches used in [13]), and the code histogram computed at each patch is concatenated to form the descriptor of the whole face image.

4. Our approach

In this section, the learning-based encoding method is transformed to face age estimation.

4.1. Patch-based code learning

In Cao et al. [13], the code set is learned using the sampled vectors from the whole face. However, the histograms are derived at the level of regions (patches). The histogram is constructed from the sampled vectors in each patch. These histograms are concatenated later to form the global descriptor.

There are variations among different face patches. Each individual patch may have different codes or code distributions, e.g. some codes may appear frequently in one patch while they are rare for another patch. To illustrate this point we build two code sets from 2080 training images (used in Section 5). One code set is learned from the sampled vectors extracted from the whole face, and the other is learned from the sampled vectors extracted from one face patch (the upper left). Later, we extracted the sampled vectors from the upper left patch in 664 testing images (also used in Section 5), then we encoded the vectors using the two code sets and constructed the frequency histograms. Fig. 1 shows the two histograms. As can be observed, for this face patch, the codes learned from the whole face are unevenly distributed (i.e., some codes rarely appear), while the codes learned from the face patch are more uniformly distributed (i.e., they are used more efficiently). Therefore, with different code set for each individual patch, the code histogram is much more informative and compact. However, learning multiple code sets introduces increase in both time and memory complexities.

4.2. Soft encoding

When encoding the input image with the learned codebook, each sampled vector (at each pixel) is assigned to the closest code. We call

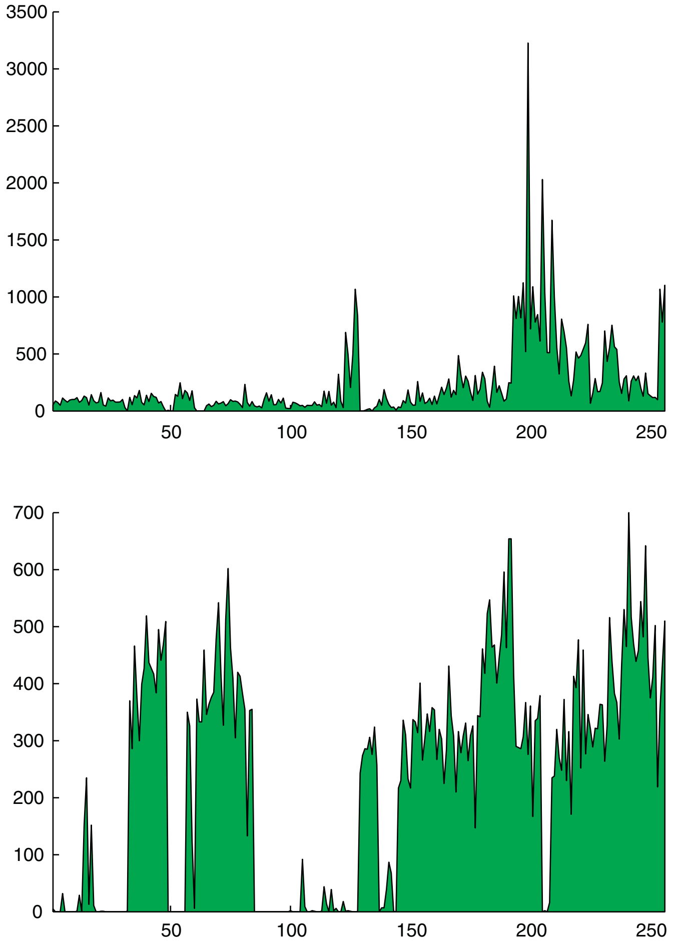

Fig. 1. The code frequency histograms in one face patch of 664 face images using two different code sets; one learned from the whole face (*top*) and the other learned from the corresponding face patch (*bottom*). 2080 face images are used for learning both code sets.

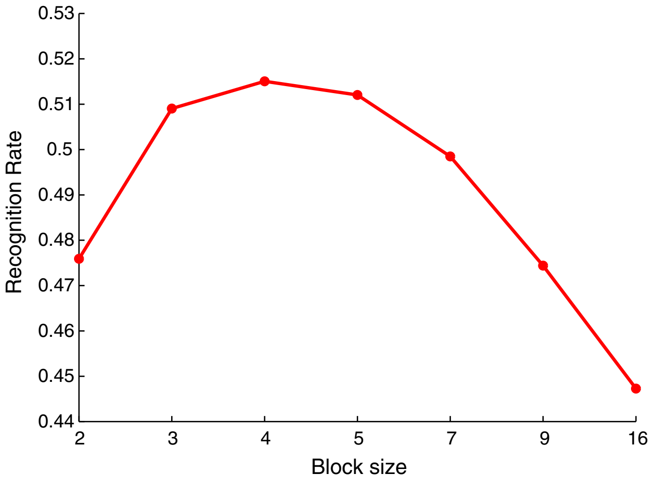

Fig. 2. The performance over different block sizes.

this hard encoding. However, for face images (especially real-world images), ambiguities always exist. That is, for a given sampled vector, there are multiple candidate codes. Assigning to the closest code makes the encoding sensitive to image noise and varying conditions (e.g. illumination). These factors can distort the sampled vector, resulting in different code assignments. We use soft encoding assigning the given sample vector to multiple codes with weights. Soft encoding is used in image classification [24].

When deriving the codes with the PCA tree, after dividing the training samples using the median value, a Gaussian distribution model is estimated for each branch. For soft assignment, the probability that it is from either branch is estimated using the Gaussian model. This is used as the weight for that branch. The weight is multiplied with the weight coming from the parent node. The new weight is passed to the children. In this way, each code (leaf) is assigned with a weight $\text{leaf}_c(r_i)$, where c is the code, and r_i is the feature vector i. The encoding is started with the weight of 1 at the tree root. The weights of all the codes are normalized. Thus the histogram bins are computed as follows:

$$Bin(c) = \sum_{i=1}^{n} \frac{\text{leaf}_c(r_i)}{S_i} \tag{1}$$

$$S_i = \sum_{c=1}^{C} \text{leaf}_c(r_i) \tag{2}$$

where C is the number of codes, n is the number of sampled vectors, and S_i is a normalization factor, i.e., the sum of weights of all codes (for the given sampled vector).

Table 1
The performance with different gradient filters.

Gradient	Result(%)	Gradient	Result(%)
Sobel1-D	51.2	Gaussian (0.5)	51.8
Sobel2-D	50.9	Gaussian (0.75)	52.8
Cubic	46.1	Gaussian (1)	52.3
Diagonal	51.7	Gaussian (3)	47.9
Prewitt	49.2	Gaussian (5)	41.9

4.3. Orientation histogram of local gradients

For each pixel, neighboring pixels are sampled in the ring-based pattern to form a low-level feature vector. However, the extracted local features are sensitive to image noise and illumination variations. Furthermore, as aging effects in faces are mainly observed as texture variations such as wrinkles and other skin artifacts, local gradients (or edge responses) may be more effective. Following HOG [7], we extract the orientation histogram of local gradients in neighborhood as the low-level feature vector for code learning.

Therefore, we use the following approach. Given a pixel, local gradients in the neighborhood (i.e., local block) are computed, and a 1-D histogram of gradient directions is accumulated over the pixels in the block. The orientation bins are evenly spaced over 0°–360°. Each gradient contributes to one or more bins, where the vote is weighted by the magnitude of the gradient; the magnitude is added to the corresponding bin. There are some parameters to choose in the implementation, including block size, gradient computing, and orientation binning. Therefore, we aim to study the influence of the various on the learning-based encoding. We use the dataset detailed in Section 5, where 1000 face images are used for code learning. 2080 training images and 664 testing images are used for age group classification using linear SVM. All faces have a resolution of 61×49 pixels. Throughout this section, results are obtained with the following default setting: 5×5 block size, 8 orientation bins (i.e., each bin covers angle of 45°), gradient computing Sobel-1D $[-1,0,1]$.

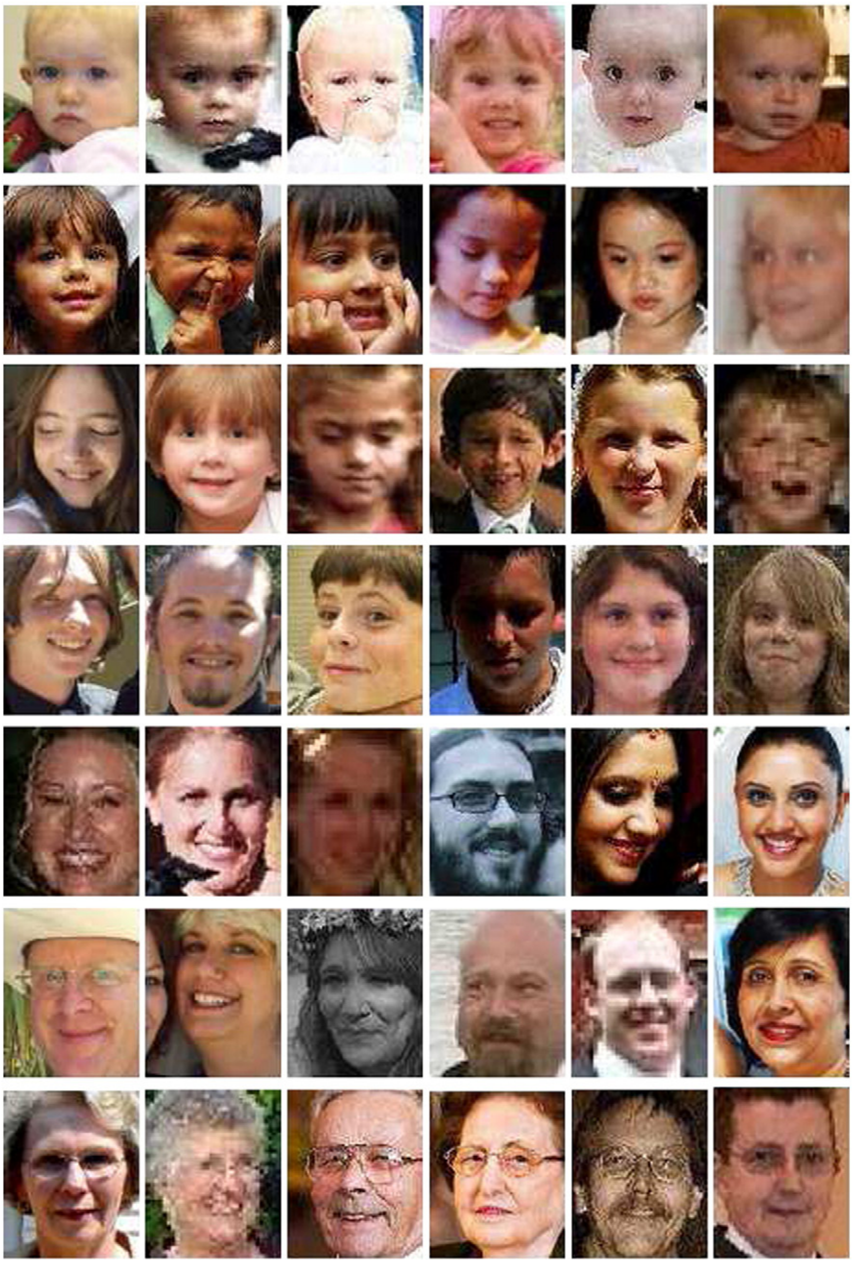

Fig. 3. Example faces in the dataset [25].

4.3.1. Block size

We test the block sizes of 2×2, 3×3, 4×4, 5×5, 7×7, 9×9, and 16×16. Fig. 2 shows the results of different block sizes when using 256 codes. It seems that the block sizes of 4×4 or 5×5 are the best choice for the dataset we use.

4.3.2. Gradient computation

We test different gradient filters, namely: Sobel-1D [-1,0,1], Sobel-2D [-1,-2,-1; 0,0,0; 1,2,1], cubic [1,-8,0,8,-1], diagonal [-1,0; 0,1], Prewitt [1,1,1; 0,0,0; -1,-1,-1] and Gaussian derivatives with different sigma values. The best performance using 256 codes is achieved using Gaussian derivatives with $\sigma = 0.75$ (Table 1). It seems that the smoothness of Gaussian helps, and fine scale derivatives perform better for this task.

4.3.3. Orientation binning

We test different bin numbers (2, 3, 4, 6, 8, 12, 16) with Gaussian and Sobel-1D gradients using 256 codes. The Gaussian derivative consistently outperforms Sobel-1D for all bin numbers. The best results are achieved using 6, 8 or 12 bins.

5. Experiments

5.1. Dataset and experimental settings

In most of the existing studies, face images with limited variations are considered. Images are usually high-quality frontal faces, occlusion-free, with clean background and limited facial expressions. However, in real-world applications (e.g. collecting demographic statistics in shops), age estimation needs to perform on real-life face images captured in unconstrained environments. There are appearance variations in real-life faces, which include facial expressions, illumination changes, head pose variations, occlusion or make-up, and poor image quality. Therefore, age estimation on real-life face images is much more challenging.

The FG-NET dataset is used in many studies. It contains face images with 68 facial landmarks. These landmarks are manually

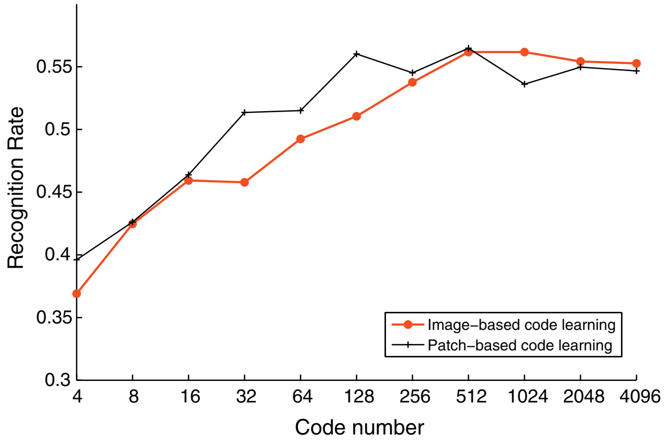

Fig. 4. The performance of image-based learning vs patch-based learning over different code numbers.

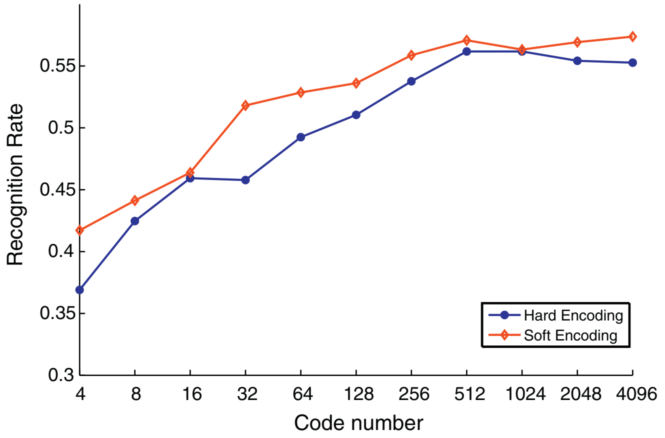

Fig. 5. The performance of soft encoding vs hard encoding over different code numbers.

detected and often used by other methods to extract shape information that helps in estimating the age [14,16,3]. However, under unconstrained conditions these landmarks cannot be accurately detected automatically. And using manually-annotated landmarks is not plausible in real-life applications. So comparing our method with other methods applied on FG-NET dataset is not feasible. To analyze the contribution of the manually annotated landmarks, Choi et al. [2] compared the performance of their method using manually and automatically obtained landmarks on FG-NET dataset. The MAE error increased around 20%.

Therefore, in this paper, we conduct experiments on real-life faces using a face image set[1] collected recently [25]. The dataset consists of 28,231 faces from 5080 Flickr images, 86% of which were detected by a face detector, and others were manually added. Each face was labeled with the gender and age category. Seven age categories were considered: 0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+, roughly corresponding to different life stages. Example faces in the dataset are shown in Fig. 3.

The dataset contains large diversity in race, pose, illumination conditions, and facial expressions. Many faces in the dataset have low resolution: the median face has only 18.5 pixels between eye centers, and 25% of the faces have under 12.5 pixels. To study age estimation on faces with reasonable resolution, Shan [22] considered only faces with the eye distance more than 24 pixels. This results in a collection of 12,080 faces. The author selected 2080 faces as the training set, and 644 faces as the testing set. The gender in the training/testing data sets is evenly distributed. In our experiments, we select another 1000 face images that are excluded from the training/testing sets for code learning, and perform age group classification using the training/testing sets. All face images are normalized to 61 × 49 pixels based on eye centers. Linear SVM is used as the classifier for simplicity. We used LIBSVM[2] for training and testing.

5.2. Experimental results

5.2.1. Code learning: image vs patches

We first examine the learning-based encoding method for age estimation. Fig. 4 shows the results. It is shown that the recognition performance increases when the code number increases for most of the code numbers. The performance decreases a bit when the code number is higher than 512. This might be due to overfitting when learning the codebook for large number of codes. The best performance of 56.2% is obtained using 512 codes. Then we compare this default image-based learning with the patch-based learning. The patch-based learning provides comparable or better performance than the image-based learning for most of the code numbers. The best performance is 56.5% with 128 codes. This suggests that code learning at the regional level leads to more informative code histogram.

5.2.2. Soft encoding

We apply soft encoding for face image encoding. The results are shown in Fig. 5. It is evident that soft encoding achieves better results than hard encoding. This illustrates that soft encoding leads to a more robust code histogram.

5.2.3. Orientation histogram of local gradients (OHLG)

We conduct experiments on code learning using the OHLG feature extraction. Based on the study in Section 4, we select the following setting: 5 × 5 block size, Gaussian derivative, and 8 orientation bins. Fig. 6 compares the results of OHLG with the sampling method. It is shown that the OHLG feature extraction produces comparable performance as the ring-based sampling. It does not outperform the sampling method. This might be due to the poor quality of the images for which the textural patterns (e.g. wrinkles) are not obvious. To verify this, we further conduct experiments on the dataset with better quality face images.

We conduct experiments on the FG-NET database [29] and MORPH database [26], both of which have better quality faces. FG-NET contains 1002 face images from Caucasian people, with the age ranging from 0 to 69 years. MORPH contains 1690 images from different ethnicities (433 Caucasian-descendant faces), with the age ranging from 15 to 68.

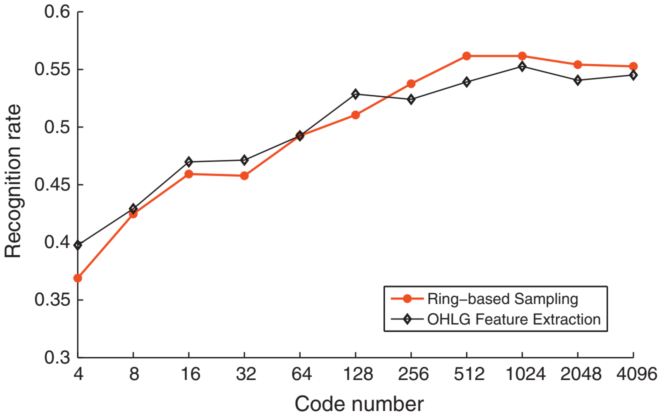

Fig. 6. The performance of code learning using the OHLG feature extraction over different code numbers.

[1] chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html.
[2] www.csie.ntu.edu.tw/cjlin/libsvm.

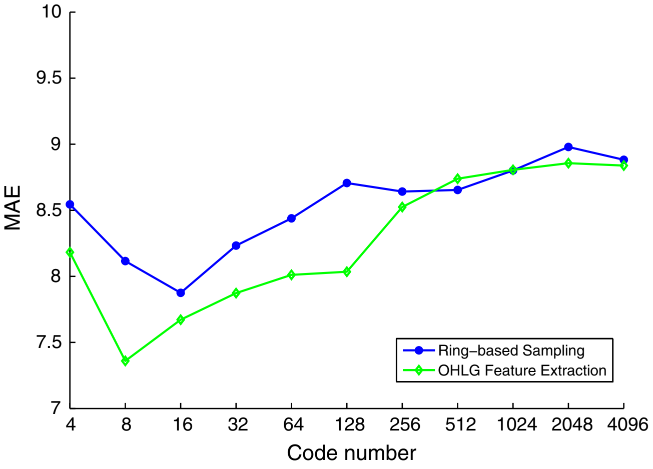

Fig. 7. The MAE on the MORPH and FG-NET dataset of code learning using the OHLG feature extraction.

We use the FG-NET data with ages between 15 and 68 as the training set, and use the 433 Caucasian images from MORPH as the testing set. The code learning is done using the remaining non-Caucasian faces in MORPH. Since we have exact ages instead of categories, we use the Mean Absolute Error (MAE) as the criterion. Fig. 7 shows the results. It can be derived that the OHLG feature extraction outperforms the sampling method in most of the code numbers. We further test soft encoding with the OHLG feature extraction on the MORPH and FG-NET dataset. The results are shown in Fig. 8. Soft encoding reduces the MAE when using the OHLG feature extraction for most of the codes, especially for larger codes. Overall, soft encoding with OHLG feature extraction outperforms the ring-based sampling for all code numbers. This illustrates the effectiveness of our improvement.

5.2.4. Codebook discriminative power

Since the codebook is learned from a separate set, the discriminative power of the images in this set and how much they reflect the differences between the age categories may affect the discriminative power of the codebook. In the following experiment, we test different sets for learning the codebook. Sets with sizes 500, 750, 1000, 1250, and 1500 are taken. The larger sets contain the smaller ones. For each set we ran the experiment using soft encoding over different code numbers. The performance is evaluated by a 2-fold cross-validation over the training set. This is to ensure that the learning code set does not fit the test set. The results are shown in Fig. 9. We noticed that the 750-image set gave the best results. This suggests that the corresponding codebook is the most discriminative. The codebooks learned from the larger sets result in lower performances. It is possible that images outside the 750-image set may contain noise negatively affecting the discriminative power of the codebook. We reran the experiment using the codebook learned from the 750-image set. Following the setup in [22], we train the descriptors over all the training set images and reported the results on the testing set in Fig. 10. The highest recognition rate of 59.5% was achieved using 1024 codes. This is 3.6 point higher than the last reported result in [22], where the recognition rate was 55.9%. The results are still not high. This is due to the variety in the images (see Section 5.1). People often have glasses, many faces are partially occluded and non-frontal and have different facial expressions. Also many images are taken outdoor with different lighting conditions.

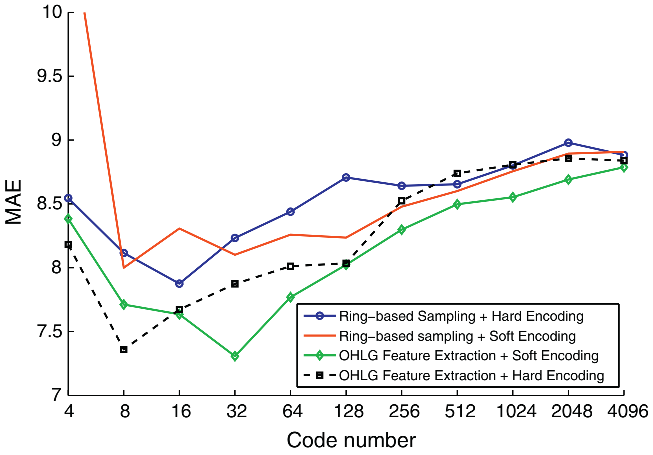

Fig. 8. The MAE on the MORPH and FG-NET dataset using soft encoding with the OHLG feature extraction.

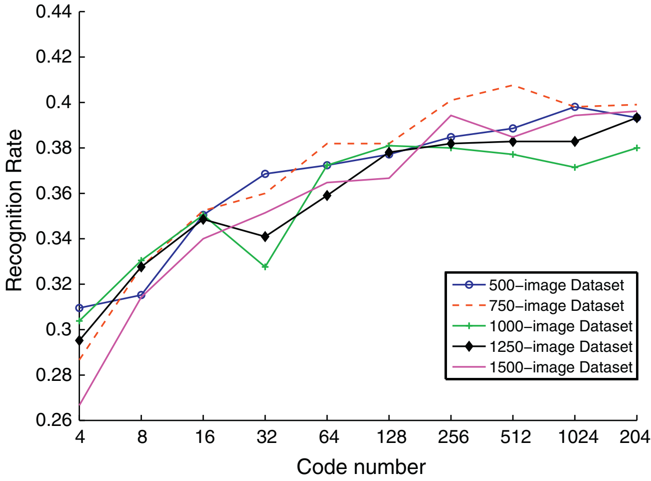

Fig. 9. Soft encoding using different learning code sets. The results were computed using two-fold cross-validation on the training set.

5.2.5. Performance on FG-NET dataset

As we explained in 5.1, the landmarks in FG-NET database are manually labeled and often used by other methods to extract shape information that helps in estimating the age. In this paper, we try to estimate the age for real life images. So we do not use the shape information of the face since, in general, no landmarker can detect the landmarks accurately enough for real-life images.

We conduct an experiment with FG-NET. The codebook is learned from a separate dataset since the FG-NET dataset has only 1002 face

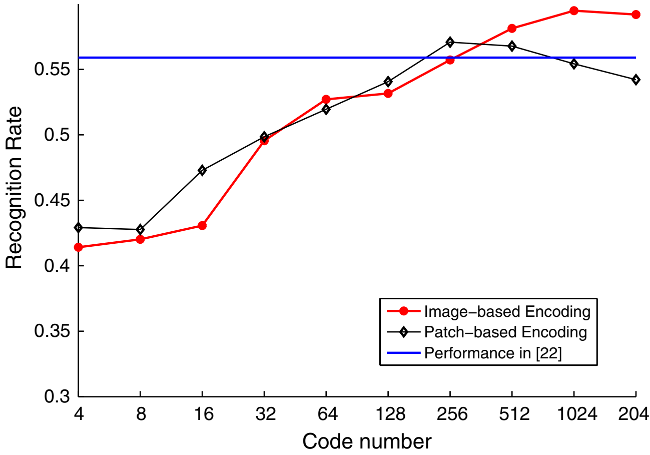

Fig. 10. Soft encoding using 750-image learning code set. The red curve represents Image-based encoding results while the black one represents patch-based encoding results.

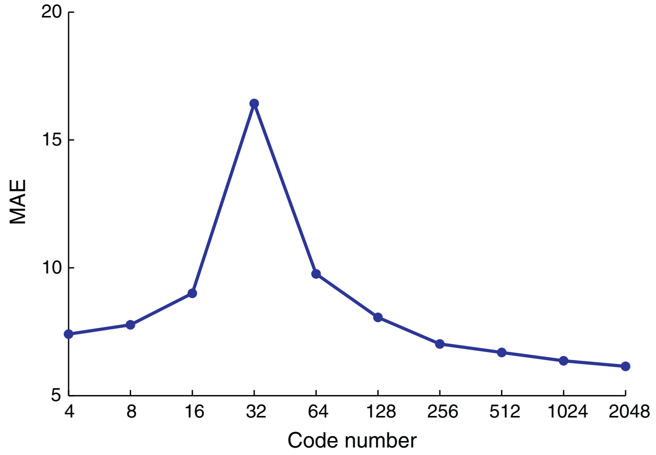

Fig. 11. The performance of code learning using soft assignment with FG-NET dataset.

images. 750 images are used for learning the codebook. The faces are cropped to 61 × 49. Then LOPO (Leave One Person Out) protocol is used for evaluating. For the estimation, we used linear support vector regression without tuning any parameters (C equals 1). We did not try different sizes of sub-datasets to learn the codebook. Fig. 11 shows the results using soft assignment.

The MAE obtained using 2048 codes is 6.14 years. The performance degraded a lot using 32 codes. But the error decreased with the increase of the code number. The lowest MAE with FG-NET and LOPO protocol was obtained by Choi et al. [2] which is 4.66 years. However, the MAE increased to 5.59 years when automatic landmarker is used.

5.2.6. Face verification

We apply soft encoding to the face verification problem. The LFW benchmark [19] is used. The LFW test set consists of 10 subsets each containing 300 same-person pairs and 300 different-persons pairs. The evaluation is reported using 10 fold cross-validation. At each fold, one subset is used for testing and other 9 are used for training. The final results are the average of the 10 fold results. Another 1000 images are used for learning the codebook. The face size is 96 × 84. As in [13], we apply a DoG preprocessing step and the codes are learned once for all the 10-folds. The 1000 image identities, used for learning the codebook, never appear in the 10 sets. Fig. 12 shows the results.

Soft encoding achieves higher results than hard encoding for most of the code numbers. This suggests that our method can be directed to other face-related problems. The reported results in [13] are around 5% higher than our results. Cao et al. used another commercial software for face alignment.[3] This might explain the difference of the results. In their paper, Cao et al. applied further dimensionality reduction and normalization steps. Here we compare with the raw feature vectors.

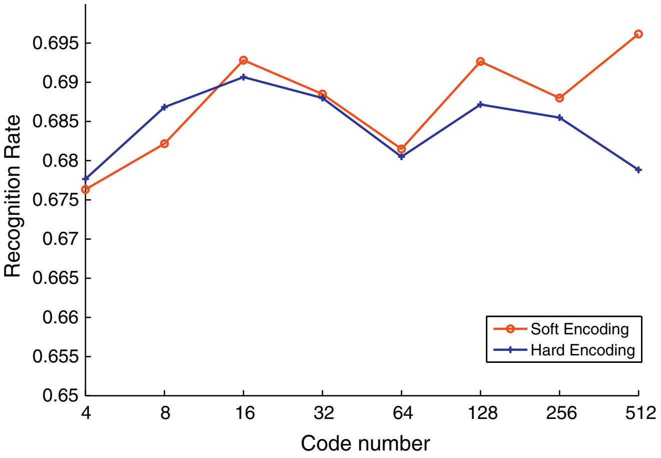

Fig. 12. The performance of soft encoding vs hard encoding over different code numbers on LFW face verification dataset.

6. Conclusions

In this paper, we adopted the learning-based encoding method for age estimation. Instead of learning a set of codes from the entire face, we extracted and learned multiple codebooks for individual face patches. Soft encoding has been used. Orientation histogram of local gradients in neighborhood has been introduced as feature vector for code learning.

Experiments showed that our extensions produced better or comparable performance for most of the cases. Using discriminative codebook, our method outperforms the best performing method reported on Gallagher dataset [22]. We extend our method to face verification and show improvements which suggests that our method can be directed to other face-related problems.

References

[1] Y. Fu, G. Guo, T.S. Huang, Age synthesis and estimation via faces: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1955–1976.
[2] S.E. Choi, Y.J. Lee, S.J. Lee, K.R. Park, J.H. Kim, Age estimation using a hierarchical classifier based on global and local facial features, Pattern Recognit. 44 (2011) 1262–1281.
[3] S. Yan, H. Wang, Y. Fu, J. Yan, X. Tang, T. Huang, N. Ramanathan, R. Chellappa, Synchronized submanifold embedding for person-independent pose estimation and beyond, IEEE Trans. Image Process. 18 (2009) 202–210.
[4] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Trans. Pattern Anal. Mach. Intell. 21 (12) (December 1999) 1357–1362.
[5] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[6] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[7] N. Dalal, B. Triggs, Histogram of oriented gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.
[8] W. Gao, H. Ai, Face gender classification on consumer images in a multiethnic environment, in: International Conference on Biometrics (ICB), 2009, pp. 169–178.
[9] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: European Conference on Computer Vision (ECCV), 2004, pp. 469–481.
[10] Z. Yang, H. Ai, Demographic classification with local binary patterns, in: International Conference on Biometrics (ICB), 2007, pp. 464–473.
[11] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vision Comput. 27 (6) (2009) 803–816.
[12] T. Gritti, C. Shan, V. Jeanne, R. Braspenning, Local features based facial expression recognition with face registration errors, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2008.
[13] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2707–2714.
[14] X. Geng, Z.-H. Zhou, K. Smith-Miles, Automatic age estimation based on facial aging patterns, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 2234–2240.
[15] Y. Fu, T.S. Huang, Human age estimation with regression on discriminative aging manifold, IEEE Trans. Multimed. 10 (2008) 578–584.
[16] G. Guo, Y. Fu, C.R. Dyer, T.S. Huang, Image-based human age estimation by manifold learning and locally adjusted robust regression, IEEE Trans. Image Process. 17 (2008) 1178–1188.
[17] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, Regression from patch-kernel, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[18] S. Yan, M. Liu, T.S. Huang, Extracting age information from locally spatially flexible patches, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2008, pp. 737–740.
[19] Gary B. Huang, Manu Ramesh, Tamara Berg, Erik Learned-miller, Faces in the wild: a database for studying face recognition in unconstrained environments, University of Massachusetts, Amherst, Technical Report, 2007.
[20] G. Guo, G. Mu, F. Yu, T.S. Huang, Human age estimation using bio-inspired features, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 112–119.
[21] B. Ni, Z. Song, S. Yan, Web image mining towards universal age estimator, in: ACM International Conference on Multimedia, 2009, pp. 85–94.
[22] C. Shan, Learning local features for age estimation on real-life faces, in: ACM Workshop on Multimodal Pervasive Video Analysis, 2010.

[3] After personal communication with the first author of the paper (Cao).

[23] Y. Freund, S. Dasgupta, M. Kabra, N. Verma, Learning the structure of manifolds using random projections, in: Advances in Neural Information Processing Systems (NIPS), 2007.

[24] J.C. van Gemert, C.J. Veenman, A.W.M. Smeulders, J.-M. Geusebroek, Visual word ambiguity, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1271–1283.

[25] A. Gallagher, T. Chen, Understanding images of groups of people, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 256–263.

[26] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2006, pp. 341–345.

[27] X. Meng, S. Shan, X. Chen, W. Gao, Local visual primitives (LVP) for face modelling and recognition, in: International Conference on Pattern Recognition, 2006.

[28] S. Xie, S. Shan, X. Chen, X. Meng, W. Gao, Learned local Gabor patterns for face representation and recognition, Signal Process. 89 (2009) 2333–2344.

[29] The FG-NET Aging Database, http://www.fgnet.rsunit.com2011.