LETTER

Common and Adapted Vocabularies for Face Verification

Shuoyan LIU[†a)], ***Nonmember and*** **Kai FANG**[†], ***Member***

SUMMARY Face verification in the presence of age progression is an important problem that has not been widely addressed. Despite appearance changes for same person due to aging, they are more similar compared to facial images from different individuals. Hence, we design common and adapted vocabularies, where common vocabulary describes contents of general population and adapted vocabulary represents specific characteristics of one of image facial pairs. And the other image is characterized with a concatenation histogram of common and adapted visual words counts, termed as "age-invariant distinctive representation". The representation describes whether the image content is best modeled by the common vocabulary or the corresponding adapted vocabulary, which is further used to accomplish the face verification. The proposed approach is tested on the FGnet dataset and a collection of real-world facial images from identification card. The experimental results demonstrate the effectiveness of the proposed method for verification of identity at a modest computational cost.

key words: *face verification, bag-of-words, concatenation histogram*

1. Introduction

Face verification across age progression has been the topic of in-depth research with widespread applications [1], [2]. It is commonly accepted that constructing an age-invariant and distinctive representation is a crucial step towards solving these problems. However, a large number of works [3]–[8] attempted to represent facial images based on the mathematical models. Lanitis et al. [3] used a statistical model to capture the variation of facial shapes as age progression. Ramanathan and Chellappa [4] applied a face growing model for face verification tasks for people under the age of 18. Biswas et al. [5] studied feature drifting model on face images among different ages and applied it to face verification tasks. Other studies [6]–[8] used age transformation model for verification. When comparing two photos, these methods either transformed one photo to have the same age as the other, or transformed both to reduce the aging effects.

Such an approach usually performed poorly since the statistical model adopted alone cannot fully describe the identity content of a facial image. This raised another face verification method which extracts features based on the cognitive knowledge. From the number of aging images, it can be seen that the same facial image is more similar to itself even after appearance changes, compared to other facial image. For this, we propose the common and adapted vocabularies to accomplish the face verification. Not much attention has been paid to using them for face verification. However, one effort in the scene categorization is proposed by Perronnin [12], which employed Gaussian Mixture Models (GMM-MAP) model to describe the universal and adapted vocabulary. Experiments have indicated that best results are obtained by adapting the mean vectors only in the verification field [2]. As an alternative, we consider maximum a posteriori vector quantization (VQ-MAP) to adapt the vocabularies using the mean vectors. In addition, the speedup originates mostly from the replacement of the Gaussian density computations with squared distance computations, leaving out the exponentiation and additional multiplications.

To be specific, the common vocabulary describes contents of general population and adapted vocabulary represents specific characteristics of one of image facial pairs. We model face verification as a two-class (intra-personal and extra-personal) classification problem. Given the age-invariant distinctive representation, the task is to assign it as either intra-personal or extra-personal using Support Vector Machine (SVM) model. The proposed approach is tested on the FGnet dataset and a collection of real-world facial images from identification card. The experimental results demonstrate the effectiveness of the proposed method for verification of identity.

2. Face Verification Using Age-Invariant Distinctive Representation

An overview of the architecture is shown in Fig. 1. The first step is to generate the common vocabulary. And then we adapt the common vocabulary based on the one of image facial pairs. The other image is characterized as the age-invariant distinctive representation. At the end of process, we first train a verification classifier V as a two-class (intrapersonal/extrapersonal) classification problem based on the age-invariant distinctive representation in the training set. For the test image pairs (D_1, D_2), we then use the classifier V to assign the appropriate label, in order to determine whether two face images D_1 and D_2 belong to the same person or different persons.

2.1 Common and Adapted Vocabularies

The common vocabulary is supposed to represent the con-

Manuscript received May 18, 2015.
Manuscript revised August 28, 2015.
Manuscript publicized September 18, 2015.

[†]The authors are with Institute of Computing Technology, China Academy of Railway Sciences, Beijing, China.
a) E-mail: 06112062@bjtu.edu.cn

DOI: 10.1587/transinf.2015EDL8117

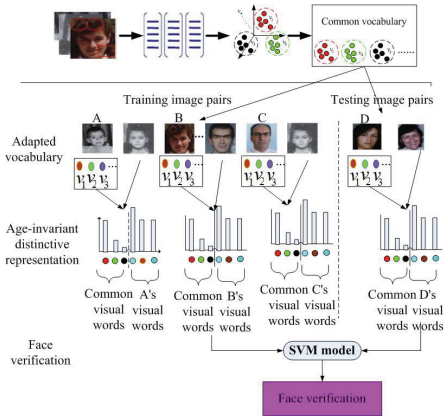

Fig. 1 Work flow of the proposed approach

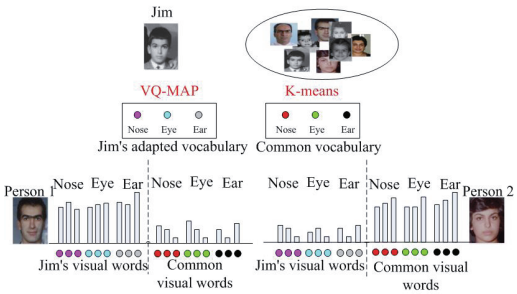

Fig. 2 In-depth analysis of age-invariant distinctive representation

tent of all possible images, and it is therefore typically trained with data from all classes under consideration. Let $X = \{x_i, i = 1, \ldots, I\}$ be the set of Scale Invariant Feature Transform (SIFT) features [9] extracted from the set of patches from training samples. Our SIFT features are designed following a great deal of work in face verification which is effective first steps toward dealing with image conditions. SIFT is invariant to image scale and rotation, and is shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination.

In special, we first use K-means to perform the clustering on all the patches, with the number of clusters provided first. Denote each cluster as a common vocabulary, we could obtain a collections of discrete common vocabularies $C = \{c_i\}_{i=1}^{n}$, where n is the number of vocabularies. The second step is to employ Maximum a Posteriori Vector Quantization (VQ-MAP) to adapt the common vocabulary using the specific characteristics of the one person. Let d_f denote the set of SIFT features extracted from the first image f in the pairs.

The adaptation centroid vector $\Theta = (a_1^f, \ldots, a_n^f)$ is performed using the following steps: Set $a_k = c_k$ for $k = 1, 2, \ldots n$; For d_f, find the index of the nearest neighbor in the adapted model $q_f = \arg\min_{1 \le k \le n} \|d_f - a_k\|^2$; For the kth cluster, define the set of vectors mapped into that cluster as $S_k = \{d_f \in X | q_f = k\}$ and calculate the centroid vector $\overline{x}_k = \frac{1}{|S_k|} \sum_{x_f \in S_k} x_f$; For the k th cluster, update the adapted vector as $a_k = \omega_k \overline{x}_k + (1 - \omega_k) c_k$, where $\omega_k = \frac{|S_k|}{|S_k|+1}$.

2.2 Age-Invariant Distinctive Representation

Once the common and adapted vocabularies have been estimated, the second image is characterized with a concatenation histogram of common and adapted visual words counts, termed as "age-invariant distinctive representation". Let d_s denote the set of SIFT feature vectors extracted from the second image s. We quantize the feature vectors d_s into one vocabulary according to the nearest neighbor rule. Thus a given image s is represented a histogram $H_s = \{c_1, \ldots c_n, a_1, \ldots, a_n\}$, where the each bin c_i is the value of i^{th} common visual word and a_j is the value of j^{th} adapted visual words.

In-depth analysis of age-invariant distinctive representation is shown in Fig. 2. Since most attributes can be shared across all people, we cluster the features into visual words to represent "eye", "ear", "nose" and etc. However the visual words adopted alone cannot fully accomplish the face verification, since age progression can cause some attributes of the same person to be different. Hence, this paper incorporates the common and adapted vocabularies to accomplish the face verification task. It can be seen that the same facial image is more similar to itself even after appearance changes, compared to other facial image. Figure 2 seems to support this, as the "eye" of person1 is more similar to the "Jim's eye" compared to the common "eye".

Our superior performance compared to the difference space [10] could be due to reserve more discriminative information. There are the main shortages to construct the difference space based on the subtraction of feature space. For example, some facial attributes (such as the eyes, nose etc.) are unconsidered when the visual features of face space with similar values. Different from them, the age-invariant distinctive representation not only reserves discriminative information, but also avoids the instability problem.

2.3 Face Verification Based on Age-Invariant Distinctive Representation

The face verification is modeled as a two-class classification problem. We first train the separating boundary, which divides the feature space into two classes (intra-personal and extra-personal pairs). We denote the separating boundary with the following equation:

$$\sum_{j=1}^{N_z} \alpha_j y_j K(z_j, H_s) + b = \Delta \tag{1}$$

where N_z is the number of support vectors and z_j is the j^{th} support vector. $K(.,.)$ is the kernel function that provides SVM with nonlinear abilities and b is a constant term. α_j is Lagrange multiplier. Each H_s is age-invariant distinctive representation. y_j is either 1 or -1, indicating the class to which the point H_s belongs. Δ is used to trade off the Recall

and the Precision.

Given an input image pair (D_1, D_2), the face verification proceeds in two stages: first, the specific feature of D_1 is used to adapt the common vocabulary. And then D_2 is characterized with a concatenation histogram of common and adapted visual words counts. Finally, the trained SVM model is used to assign intra-personal or extra-personal label to D_2 based on the concatenation histogram.

3. Experiment Results

We start our experiments with the in-depth analysis of our method on the FGnet dataset [11]. The FGnet Aging Database is widely used for research of age-related facial image analysis. Some examples of the FGnet dataset are shown in Fig. 3. For verification tasks, we generate 665 intra-personal pairs by collecting all image pairs from the same subjects. Extra-personal pairs are randomly selected from images from different subjects.

Two numeric performance measures often considered are the Recall, defined as the proportion of positive cases that are correctly identified, and the Precision, defined as the proportion of the predicted positive cases that are correct. The average recall and precision rate of the proposed approach is approximately 92.2% and 93.2%, respectively. To study the aging effects on face verification, we further divide the 665 intra-pairs into four different age-gaps (0-2 years, 3-4 years, 5- 7 years, 8- 10 years), which is summarized in Table 1. To be specific, we construct the 225 intra pair in the 0-2 years, 154 intra pair in the 3-4 years, 121 intra pair in the 5-7 years and 165 intra pair in the 8-10 years, respectively. It can be seen that the mean of age difference is in the range of corresponding age gap.

The intra-personal image pairs are further classified into the above four age-difference categories in the Fig. 4. In the confusion table, the x-axis represents the results of the proposed approach. The y-axis represents the ground truth. Hence, main diagonal line gives the proportion of samples correctly predicted. Data on non-main diagonal lines indicates the proportion of the sample which is not correctly predicted. It can be seen that the age-invariant distinctive representation can indeed accomplish the face verification.

Fig. 3 Example images from the FGnet dataset

Table 1 The number of pairs in the four age gap

Age gap	0-2years	3-4years	5-7years	8-10years
#intra-pair	225	154	121	165
mean age diff.	1	3.5	6.2	8.6

A closer look at the confusion table reveals that the higher error occurs in the 8-10 years age gap. It may decrease the power of the proposed representation to handle larger age gap.

We subsequently investigate how the verification performance is affected by the number of visual words. Figure 5 shows this performance variation for the four differently sized visual words 10, 20, 30, 40 and 50. It can be seen that the performance increases progressively until visual words size is 30, and then drops off slightly. It demonstrates that if the number of visual words is too small, it is easily to sacrifice the discriminative power of the vocabulary. When the number of categories is large, it makes the histogram computation costly, and it makes the classification of the histograms particularly challenging, especially in the case of scarce training data, as the dimensionality of the histograms may become far greater than the number of available samples. Through comparing, here the visual words number is set as 30 in experiments, which is enough for modeling various age gaps of facial appearance.

Figure 6 compares verification rate of the proposed method with [4] on FGnet dataset. From the Fig. 6, the proposed approach is better than the approach of Ramanthan [4] in all four age categories. The paper [4] constructed the difference space by the subtraction of bin values of histogram. This setting discards the bins when their corresponding visual words with similar values. Nevertheless, each bin in the histogram corresponds to a facial attribute (such as the eyes, nose etc.). In addition, Ramanthan [4] assumed the face difference space meets Gaussian distribution, since not all samples fit to Gaussian distribution.

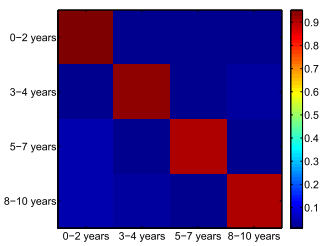

Fig. 4 The confusion table based on the age-invariant distinctive representation in the four age gaps

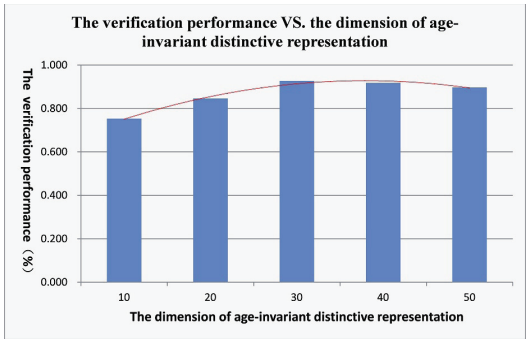

Fig. 5 The verification performance of various visual words

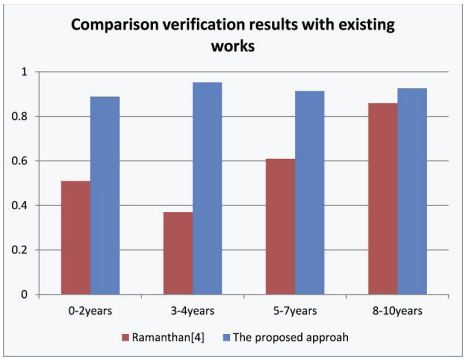

Fig. 6 Comparison verification results with existing works

Train ticket real-name system is the important application for the face verification. We construct a new data set of real-world image pairs acquired from the identification card and mugshot images. To evaluate the validity of the proposed face verification algorithm in the railway ticket real-name system, we use all of the FGnet images as training data, and test the proposed approach on the real-world image pairs. The average performance of this approach achieves 86.6%. In addition, we now provide a breakdown of the computational cost, since our emphasis in this work is on practicality. To estimate the cost, we run our code on a PC with Intel Core i7 with 3.4GHz and 8 GB of RAM. The base cost of the verification process on a new pair of input images is approximately 400 ms. This may be split into 150 ms for the SIFT feature extraction and 250 ms for the adapted vocabulary construction. The additional cost is approximately 5ms for computation of age-invariant distinctive representation and the classification step. The proposed approach is generalizable, as they can be learned once and then applied to verify identity of image pairs without any further training. Hence, once age-invariant distinctive representation has been computed, the computation of the verification process is negligible. This approach is very practical since it has reasonable computational costs.

4. Conclusion

This paper proposes the common and adapted vocabularies to tackle the face verification. The experimental results demonstrate the effectiveness of the proposed method for verification of identity at a modest computational cost. However, the task of face verification still leaves room for improvement. This is a promising direction, and we might be able to for much more powerful representation for face verification.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition," A literature survey, ACM Computer Survey, vol.35, no.4, pp.399–458, 2003.
[2] Y. Fei, Face Recognition Technology for confirmation task [D], Graduate School of the Chinese Academy of Sciences, 2006.
[3] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Toward automatic simulation of aging effects on face images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.24, no.4, pp.442–455, 2002.
[4] N. Ramanathan and R. Chellappa, "Face verification across age progression," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.462–469, 2005.
[5] S. Biswas, G. Aggarwal, N. Ramanathan, and R. Chellappa, "A non-generative approach for face recognition across aging," Proc. IEEE Conf. Biometrics: Theory, Applications and Systems, pp.1–6, 2008.
[6] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic Age Estimation Based on Facial Aging Patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.12, no.29, pp.2234–2240, 2007.
[7] J. Suo, X. Chen, S. Shan, and W. Gao, "Learning long term face aging patterns from partially dense aging databases," Proc. Int. Conf. Computer Vision (ICCV'09), pp.622–629, 2009.
[8] U. Park, Y. Tong, and A.K. Jain, "Age-Invariant Face Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.32, no.5, pp.947–954, 2010.
[9] D. Lowe, "Distinctive image features from scale-invariant key points," International Journal of Computer Vision, vol.60, no.2, pp.91–110, 2004.
[10] H. Ling, S. Soatto, N. Ramanathan, and D.W. Jacobs, "Face verification Across Age Progression Using Discriminative Methods," IEEE Trans. Information Forensics and Security, vol.1, no.5, pp.82–91, 2010.
[11] The FG-NET Aging Database 2002 [Onlin]. Available: http://grail.cs.washington.edu/aging/FGNET.zip
[12] F. Perronnin, "Universal and adapted vocabularies for Generic visual Categorization," IEEE Trans. PAMI, vol.30, no.7, pp.1243–1256, 2008.