Video Face Recognition based on External General Dictionary

Qing Liu, Shaohu Peng, Jiadong Wang, Xiao Hu[a]

School of mechanical and electric engineering, Guangzhou University, Guangzhou,510006, China

[a]Email: huxiao@gzhu.edu.cn

Keywords: Face recognition; generic variation dictionary; sparse representation

Abstract. Face recognition with single sample per person (SSPP) was a very challenging task because SSPP cannot reflect the characteristics of the test samples (eg. illuminations, pose, expressions and disguises) so that it is difficult to predict the facial variations of a query sample by the gallery samples. Considering the fact that the intra-class facial variations can be shared across different subjects, a novel algorithm was put forward. This algorithm took generic variation dictionary to solve the problem of lack the number of sample. This method need create two different dictionary. Generic variation dictionary *D* was built by the different group of pictures and gallery dictionary *X* was built from the training samples. Using half-quadratic optimization function to solve optimization problem. The experimental results showed that the recognition rate of the algorithm was superior to recognition rate of some traditional methods.

Introduction

Face Recognition(FR) is an important study direction of computer vision and could be applied in many aspects[1].Although the FR has been studying for many years there still exists so much challenge .For example, capturing the original image from the video will be influenced by illuminations, pose, expressions and disguises. Even in actual application only has one sample image for machine training for the duration of our experiment. But most of the face recognition techniques will suffer serious performance drop when there is only one training sample per person[2]. For solving this problem[3] advocate the method of 3D technology to build virtual sample to solve the deficiency of training sample . In [4] adopt sample expansion, adding some virtual samples that correspond with reality and it is of great importance for the heightening of final discrimination. In [5] takes advantage of SVD division All methods above mentioned had increased face recognition rate, but the result is not obvious enough The major reason is that the virtual sample generated by those method has high relevance to the original image. This sample cannot be considered as independent samples for feature extraction. In[6] put forward a new method with constructing an external versatile train set to resolve the problems bring with by the deficiency of sample. But when constructing the external dictionary, the static database they adopted has obvious feature It is so difficult to acquire so obvious feature of image as an external general dictionary training sample in practice. .In this paper, we propose a new learning method of external general dictionary. The training sample of external general dictionary obtained by the camera with different distance, times and angles is similar. get numerous set of samples will be able to form a reference transformation can be set to training into a dictionary. Acquiring dictionary by this way not only full of representative feature transform, but also is practicable

Face Recognition based on External General Dictionaries

According to the theory of Sparse Representation Classification (SRC) [7] (for the given signals (image) $z \in R^{m\times 1}$ .If there is sufficient number of representative training samples. It could be linear description by the dictionary $A=(a_1,a_2,...,a_n)$ .where $a_k \in R^{m\times 1}$, $k=1,2,...,n$. However for FR with single sample per person(SSPP), we have a gallery set $X=[x_1,...,x_k,...,x_K]\in R^{d\times k}$, where $x_k \in R^d$ is the only single gallery sample of class k, k = 1, 2, ...,K. Given a query sample $z \in R^d$,

representation based classifiers such as SRC represent it over the gallery set X as:

$$z = Xa + e \quad (1)$$

According to the theory of compression sensing [7] . If the gallery set has many training samples for each subject, most of the facial variations in the query sample can be synthesized by the multiple samples from the same class, and consequently correct classification can be made via comparing the representation residual of each class. For FR with SSPP, unfortunately, there is only one training sample of every subject. However for most of the algorithms the recognition rate will fall sharply.As the intra-class facial variations caused by illuminations, pose, expressions, disguises and these cases can be shared between the images, an external generic training set which consists of numerous face images with various types of variations can be adopted to construct an intra-class variation dictionary D . So even in the case of small sample can also be a good reflection of the characteristics of the face, which will make up for the lack of internal characteristics of the face of the lack of sample. The improved sparse expression as[6]:

$$z = Xa + D\beta + e \quad (2)$$

The dictionary D is acquired by the set external training and the training process is $G = [G^r - G^v]$ where G^r and G^v are the reference subset and variation subset, respectively. The reference subset $G^r \in R^{d \times n}$ is composed of neutral face images or the mean faces of each subject. The variation subset G^V involves M possible facial variations: $G^v = [G_1^v, ...G_m^v, ...G_M^v]$, where G_m^v is the subset of the m[th] variation, m =1, 2, ...,M. In [3], a sparse variation dictionary is learned from G. In our work,we simply construct an intra-class variation dictionary, denoted by D, by using

the difference between G^r and G^v :

$$D = [G_1^v - G^r, ..., G_m^v - G^r, ...G_M^v - G^r] \in R^{d \times nM} \quad (3)$$

Considering the practicality of video , we can take the face images from different distance ,different time and different angle images as the training sample. Because these images are not the same. With these differences of the image, we can construct the external general dictionary though training.

(a)original image (b)cam-1 image (c)cam-6 image

Fig. 1. reference image and feature transform image

As is shown in Fig.1. Fig.1 (a) as a reference set of positive HD photos. Fig.1 (b) and (c) respectively for camera distance of 1 meters of camera in obtained photographs and infrared camera in the evening made photos. It could be considered as the reference subset of the expression and illumination respectively . With these differences, we can construct the external general transform dictionary .

In this paper we know the corresponding expression can be written as:

$$z = Xa + D\beta + e \quad (4)$$

where α and β are the representation vectors of z over X and D, respectively and e is the representation residual .In order to find out the optimal solution. Pengfei Zhu propose a new algorithm[6] that can be found through the conversion of appropriate e and the regularization matrix $R(\alpha, \beta)$.We consider the following optimization problem to solve $R(\alpha, \beta)$:

$$\min\{\alpha, \beta\} \sum_{i=1}^{S} l(\| e \|) + \lambda R(\alpha, \beta) \quad (5)$$

$$s.t. z_i = X\alpha + D\beta + \varepsilon, i = 1, 2, ...S$$

The problem now turns to how to define the loss function $e = \| e \|_2$ and regularizer $R(\alpha, \beta)$. The expression is converted as follow:

$$\{\alpha,\beta\}=\min\{\alpha,\beta\}\| z-X\alpha-D\beta\|_2^2+\lambda(\| a\|_2^2+\|\beta\|_2^2) \quad (6)$$

$$e=\| z-X\hat{\alpha}+D\hat{\beta}\|_2$$

In [9]. It is proved for the face images the distribution of e is highly non-Gaussian. In [10], the concept of correntropy is proposed to measure the loss of non-Gaussian data. A correntropy induced metric (CIM) for residual e is defined as :

$$CIM(e)=\left(k_\delta(0)-k_\delta(x)\right)^{\frac{1}{2}} \quad (7)$$

Where $k_\delta(\cdot)$ is the kernel function [11]. The kernel function $k_\delta(x)=\exp(-x^2/2\delta^2)$ is proved to be have a good robust in the representation of the image. For the $R(\alpha,\beta)$. In[12] we define it as the l_2-norm of a and β. Compared with the l_1-norm, the l_2-norm not only has a good representation but also reduces the computational cost .Finally, the representation model becomes:

$$\min\{\alpha,\beta\}\sum_{i=1}^{S}(1-k_\delta(\|\varepsilon\|_2))+\lambda(\|\alpha\|_2^2+\|\beta\|_2^2) \quad (8)$$

$$s.t.z_i=X\alpha+D\beta+\varepsilon, i=1,2,...,S$$

Our classification principle is to check which class can lead to the minimal residual .Firstly, structure a dictionary $X=[X^1,...X^k,...X^K]$ by the set of training, and the a^k is the kind of *k*. According to the theory of compression sensing, we can get a set of sparse coefficient vectors α. Secondly, it can be written as $a=[a^1;...a^k;...;a^K]$ Which a^k presents the kind of D. We calculate the sparse coefficients for each of the classes by dictionary of X^k and D .Then we can calculate the remaining coefficients of each label by *w* [6](The *w* corresponds to the each image) Finally, the query sample z is classified to the class which has the minimal weighted representation residual .

Experimental analysis

This paper extracts data from the SCface video facial database [14] for testing. The database has 4,160 pictures taken from 130 people respectively at three different distances: Distance3, Distance2, Distance1; and by 7 different cameras. Therein, these three distances are 1m, 2.6m and 4.2m respectively. Cam1~Cam7 refer to the facial images taken from 7 different cameras, among which Cam6 and 7 are taken by infrared cameras at night in order to attest the competency of the methodology in this paper. The calculation in this paper, the SCR part [15] based on single sample and the 3D method (PCA part) [16] will be compared for analysis

Derived from the total 130 people from SCace database, the first 20 people will be sampled to make up an generic variation dictionary. 40 People with serial number 21 to 60 will become training and testing samples. In the Dem1 experiment, three tests are taken in Distance3, Distance2 and Distance1 respectively. All three tests in Dem1 will adopt the same external universal exchange dictionary. The pictures contained in the dictionary are produced at different distances and by different cameras. Whiles, the three tests in Dem2 use different generic variation dictionary. The three different dictionaries are made up by pictures respectively from Distance3, Distance2 and Distance1 and different cameras. All sampling pictures will be converted to size 75*75.

Test results

The experiment uses 3D and SRC calculation methods in separate. In addition, the dictionary composed by different distances and different cameras (Dem1) and the generic variation dictionary (Dem2) made by same distance but different cameras adopt the other two calculation methods. All these calculation methods form 4 sets of experiments from three different distances in SCface video facial database. The results from the three tests are mean values as shown in Table 2.

Table 1. Recognition rate (%) on SCface database.

Method	Number	Distance3	Distance2	Distance1
3D	40	13%	19%	6.5%
SRC	40	9.6%	12%	7.2%
Dem1	40	15%	22%	10%
Dem2	40	20%	25%	12.5%

From Table 1, we can see among the 4 sets Dem2 has the best recognition result as using generic variation dictionary method as set in this paper. However, Dem1 is slightly worse than Dem2 in recognition capacity due to the fact that the pictures taken in Dem2 are from the same distances, the recognition capacity is better this way and the pictures look more similar to the originals. This brings up a very good revelation that we could possibly carry on our researches based on the existing samples when making another external exchange dictionaries in the future. But comparing with the traditional SRC method and the PCA method after the dictionary samples formed by 3D technology, the recognition capacity in this method has improved.

Conclusion

On the account of insufficient sample quantity, the training samples cannot manifest internal feature changes in this paper. Therefore, an generic variation dictionary is made to reflect the internal feature changes. Such measure not only solves the low recognition problem due to small sample amount, but also saves application operating hours and internal memory usage by sharing the same external dictionaries for all samples. This puts forth a brilliant idea to solve the video facial recognition challenge from a single sample. As can be seen from the above, the recognition result by the introduced method in this paper turns out to be better than that by the existing method. As the pictures extracted from videos are not as clear as those taken from still objects for external dictionary training, how to derive a relatively stable external dictionary from lower recognition samples becomes our next working subject.

Acknowledgement

In this paper, the research was sponsored by the Nature Science Foundation of Guangdong Province (Project No. S2013010013511) and Science and technology planning project in Guangzhou (Project No. 2014J4100127).

References

[1] Yajin Wu, Shuishen Zhou . Research On Face Recognition Method Based On Compressed Sensing Theory [D], Xi'an University of Science and Technology, 2014:7-8.

[2] Yaying Zhao, Xiaohu Ma. Research on Algorithm of Face Recognition with Single Sample per Person [D]. Suzhou University, 2012:18-19.

[3] M.D.Levine, Y.F.Yu. State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person [J].Patten Recognition Letters,2009,30:908-913.

[4] Yongjun Liu , Jingyi Chang, Caikou Cheng. Face Recognition with the single training sample per peopel based on bit-planes images and 2DMSLDA [J].Computer Engineering and Applications. 2010,46(15):172-175.

[5] Shengliang Zhang , Fubing Cheng , Jingyu Yang. Some Researches for Face Recognition with

One Training Image per People[J]. Computer Science.2006,32(2):225-229.

[6] Pengfei Zhu, Meng Yang ,Lei Zhang and Il-Yong Lee. Local Generic Representation for Face Recognition with Single Sample per Person[C].In ACCV 2014:1-16.

[7] Peng Peng, Huangrong Hao. Research on Face Recognition System Based on Compressed Sensing [D] Donghua University , 2015:14-30.

[8] Dang Liu, Junying Gan. Application of block SRC algorithm in occlude face recognition[D] Wuyi university . 2013:34-41.

[9]] Lu, C., Tang, J., Lin, M., Lin, L., Yan, S., Lin, Z. Correntropy induced l2 graph for robust subspace clustering[C]. In: Proc. 14th IEEE International Conf. Computer Vision (ICCV). (2013).

[10] Liu, W., Pokharel, P.P., Pr´ıncipe, J.C. Correntropy: properties and applications in non-gaussian signal processing[J]. IEEE International Conference on Robotics and Automation, 2007,55:5286–5298.

[11] Nikolova, M., Ng, M.K. Analysis of half-quadratic minimization methods for signal and image recovery[J]. SIAM Journal on Scientific computing, 2005,27: 937–966

[12] Zhang,L.,Yang,M., Feng,X. Sparse representation or collaborative representation; Which helps face recognition?[C]. Global Electronics. 2011:18-27

[13] Mislav Grgic, Kresimir Delac and Sonja Grgic. SCface–surveillance cameras face database[J], Multimed Tools Appl, 2011,51:863-879.

[14] Xiao Hu, Shaohu Peng, Jiyong Yan. Low Spatial Resolution Face Recognition Based on Compressive Sensing[C], International Conference on Automation, Mechanical Control and Computational Engineering, 2015:591-596.

[15] Xiao Hu, Qixin Liao and Shaohu Peng, Video Surveillance Face Recognition by Move Virtual Training Samples Based on 3D modeling[C]. 2015 11th international Conference on Natural Computation,2015: 113-117.