LETTER *Special Section on Recent Advances in Multimedia Signal Processing Techniques and Applications*

Registration Method of Sparse Representation Classification Method

Jing WANG[†a)], *Student Member and* Guangda SU[†b)], *Nonmember*

SUMMARY Sparse representation based classification (SRC) has emerged as a new paradigm for solving face recognition problems. Further research found that the main limitation of SRC is the assumption of pixel-accurate alignment between the test image and the training set. A. Wagner used a series of linear programs that iteratively minimize the sparsity of the registration error. In this paper, we propose another face registration method called three-point positioning method. Experiments show that our proposed method achieves better performance.

key words: *classification, face recognition, face registration, sparse representation*

1. Introduction

Face recognition has gained much attention in the last two decades due to increasing demand in security and law enforcement applications. Recently, using the sparse representation classification (SRC) method has attracted a lot of attention; especially after its effective application in solving the face recognition problem with significant illumination and expression variations [1]. Based on the theory of compressive sampling [2], this method exploits the discriminative nature of the sparse representation to perform classification. A lot of researches have been done and have achieved great progress [3]–[5]. SRC algorithm has achieved satisfactory recognition results on public face databases including the Extended Yale B face database [6], the AR database [7] and the CMU Multi-PIE face database [8].

However, as is pointed in [5], while those works achieved impressive results on public datasets taken under controlled laboratory conditions, it fails to address two critical aspects of real world face recognition: significant variations in both the image domain and in the image value. To solve this problem, Andrew Wagner [5] proposed a series of linear programs that iteratively minimize the sparsity of the registration error and demonstrated its validity (In this paper, we use LP to represent this method). Although this method can improve the registration accuracy effectively, it is time consuming due to the high complexity.

In this paper, we propose "three-point positioning" (TPP) method to register the face images. Eyeballs and mandibular point are used to normalize the face images; then, the SRC algorithm will be performed on these nor-

malized images. TPP method can perform registration accurately and quickly. Experiments shows that TPP achieves satisfactory performance, which is much better than LP.

The rest of this paper is organized as follows: In Sect. 2, previous related works are introduced. Our proposed method is elaborated in Sect. 3. Section 4 presents the experimental results. Section 5 concludes our work.

2. Previous Works

SRC algorithm was proposed in [1]. Through proper selection of the training samples and the number of features, the SRC algorithm can achieve satisfactory recognition results despite of serious unfavorable illumination condition or large expression variation. The fundamental assumption of this method is that the training samples from a single class lie on a linear subspace. By defining the training matrix A, the linear representation of y can be written as:

$$y = Ax_0 + \varepsilon \tag{1}$$

where $x_0 = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0]^T \in \mathfrak{R}^n$ is an unknown coefficient vector whose entries should be zero except those associated with the i-th class, and ε is a noise term.

Equation (1) can be efficiently solved through solving the l^1-minimization problem shown in Eq. (2).

$$(l^1) : \hat{x}_1 = \arg \min \|x\|_1 \text{ subject to } y = Ax + \varepsilon \tag{2}$$

Future research found that the main limitation of SRC is the assumption of pixel-accurate alignment between the test image and the training set. This leads to brittleness under pose and misalignment, making it inappropriate for deployment outside a laboratory setting [5]. They also show that this problem can be solved by a series of linear programs that iteratively minimize the sparsity of the registration error (LP). They represented warped image as $y = y_0 \circ \tau^{-1}$, for some transformation $\tau \in T$, where T is a finite-dimensional group of transformations acting on the image domain. Thus Eq. (2) can be written as:

$$\hat{\tau} = \arg \min \|x\|_1 + \|e\|_1 \text{ subject to } y \circ \tau = Ax + \varepsilon \tag{3}$$

Considering that Eq. (3) a difficult non-convex optimization problem, the following registration algorithm was used.
Algorithm 1.
  for each subject k,

$\tau^{(0)} \leftarrow I.$
do
$\quad \tilde{y}(\tau) \leftarrow \frac{y \circ \tau}{\|y \circ \tau\|_2}; \ J \leftarrow \frac{\partial}{\partial \tau} \tilde{y}(\tau) \big|_{\tau(t)};$
$\quad \tau = \arg\min \|e\|_1 \ \text{s.t.} \ \tilde{y} + J\tau = A_k x + e, \ x \geq 0.$
$\quad \tau^{(i+1)} \leftarrow \tau^{(i)} + \tau;$
while $\left\| \tau^{(i+1)} - \tau^{(i)} \right\| \geq \varepsilon$
end

Although it leads to an effective alignment algorithm for face images, the computational complexity is very expensive because of the additional iteration process.

3. Three-Point Positioning (TPP) Method

As is shown in Fig. 1, we choose eyeballs and mandibular point as the three key-points. The accurate face and eye locating method is presented in [9], [10]. However, locating the mandibular point is more challenging, due to the smaller gradient change and greater external interference (e.g. mustache and collar). So far, few methods can achieve satisfactory results in locating the mandibular point while it is essential to face recognition. As a result, we propose a mandibular point locating algorithm, which significantly improves the locating performance and achieves a satisfactory accuracy.

The symmetry of human face helps us find the x-coordinate of mandibular point easily, when the two eyeballs are located accurately. In contrary, locating y-coordinate of mandibular point is more difficult and needs more effort. When locating a mandibular point from a face image manually, we always make the judgment globally instead of locally. Therefore, it is more reasonable to use integral projection of the whole facial profile to search the mandibular point. Facial profile can be fitted with a high order polynomial curve. However, the higher order the polynomial is,

the higher the computational complexity will be. Besides, the occlusion of hair brings more variety that may interfere with the subsequent judgments. Through a large number of observations, we found that the shape of chin is similar to a parabola, and the quadratic polynomial fitting of the chin can achieve a satisfactory accuracy, as is shown in Fig. 2.

The entire parabola fitting procedure is outlined in Algorithm 2. The input to this procedure is the normalized face image I and the location of eyes (*lefteye*, *righteye*).

Algorithm 2. Parabola Fitting Procedure

1. Estimate the location of the initial mandibular point from the location of eyeballs;

2. Set mandibular as the parabolic vertex, initial value, adjusting $P(1)$ and $P(2)$ to find the matching parabola. Record the sum gradient G of each point on the parabola.

$$y - mandibular.y = P(1) \times (x - mandibular.x)^2$$
$$+ P(2) \times (x - mandibular.x)$$

3. Change the parabolic vertex within a certain range, and find the matching parabola as the step 2.

4. Arrange the 'G's in a vector, the best fitting parabola corresponds to the maximum point of the vector. The parabolic vertex is the required mandibular point.

The searching procedure of Algorithm 2 is also described in Fig. 3 (a). Through changing the y-coordinate of parabolic vertex, matching parabola is obtained step by step. After comparing the gradient of each matching parabola, the best one is marked in white color.

From the analysis above, mandibular point can be approximately determined as the vertex of the parabola that best fit the chin. However, in some cases, such a locating method may produce errors, since the parabola is a rough estimation so that pix-level accuracy cannot be achieved. Therefore, further smaller-scale search is needed when the parabola vertex is calculated, as is shown in Fig. 3 (b).

Fig. 1 The three key-points.

Fig. 2 Parabolic approximation of the chin.
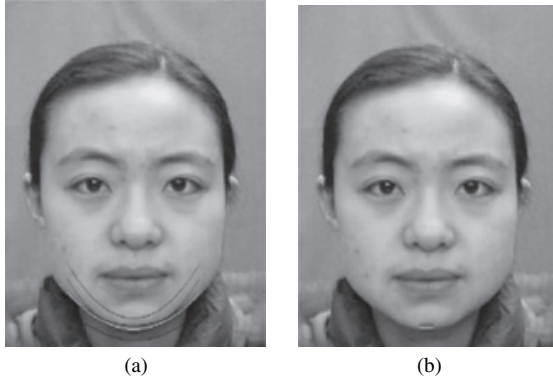
Fig. 3 Mandibular point searching procedure. (a) The first searching procedure: Parabola Fitting. (b) The second searching procedure.

Y-coordinate of mandibular point is changed in a small range, and the sum gradient of horizontal line through each mandibular point is recorded. The best mandibular point corresponds to the horizontal line that with the maximum gradient, as is marked in white color.

The first searching procedure helps to find the matching parabola robustly. The parabola can be positioned accurately in moderate illumination and expression variations. In the second searching procedure, the mandibular point is regulated in a small range to make the result more precise.

After that, the face image is normalized according to the three points. The normalization steps are: first, rotating the image until the two eyes are at a horizontal position; then, zooming the image until the vertical distance between the eyes and mandibular is 200; after that, cropping the face image so that the x-y coordinates of the mandibular point is 180–420 and the image size is 360×480.

4. Experiment Results

To verify the performance of the proposed method, we perform two sets of experiments. In the first set, we compare our mandibular point locating algorithm with ASM [12], [13] and ASM+AAM [14] on TH-FACE database. In the second set, we compare the recognition accuracies of TPP, LP and the accurately registered images on both the AR face database and the CAS-PEAL face database.

4.1 Mandibular Point Locating Results

The performance of mandibular point locating algorithm is experimented on our own face database (TH-FACE). TH-FACE contains five thousand of frontal face images under well-controlled lighting condition. 1,000 face images were chosen to perform the experiment. We also applied the ASM and ASM+AAM algorithm to locate the mandibular point as a comparison. Experimental results are shown in Table 1.

In Table 1, $err \leq 3$ refers to the proportion of point i among the 1,000 test images, whose locating deviation is less than C pixels. Our method is more accurate than ASM and ASM+AAM algorithm.

Table 1 Mandibular point locating results.

Method	$err \leq 3$	$err \leq 6$	$err \leq 9$
ASM method	0.948	0.992	0.996
ASM+AAM method	0.959	0.992	0.996
Our method	0.970	0.993	0.998

Table 2 Recognition accuracies on AR face database.

Feature Length	LP (%)	TPP (%)	Accurately registered images (%)
30	56.78	74.92	80.03
130	61.59	84.76	90.00
300	80.32	90.63	91.59
560	80.11	90.95	92.22

Table 3 Recognition accuracies on CAS-PEAL face database.

Feature Length	LP (%)	TPP (%)	Accurately registered images (%)
30	74.01	78.71	80.20
56	83.91	89.85	91.58
168	89.11	91.34	93.56
700	89.85	91.34	92.08

4.2 Face Recognition Results

In this subsection, we do experiments on public face databases to evaluate the performance of the two registration method. We also compare the performance with the accurately registered images. Notice that the choice of features is no longer critical when sparsity in the recognition problem is properly harnessed [1], downsampled images are used during the experiments.

The AR database consists of over 4,000 frontal images for 126 individuals. For each individual, 26 pictures were taken in two separate sessions [7]. These images include facial variations including illumination change, expressions, and facial disguises. In the experiment, we choose a subset of the dataset consisting of 45 male subjects and 45 female subjects. For each subject, 14 images with only illumination change and expressions were selected: the seven images from Session 1 for training, and the other seven from Session 2 for testing. The images are all converted to grayscale.

The CAS-PEAL face database contains 99,594 images of 1,040 individuals (595 males and 445 females) with varying pose, expression, accessory, and lighting (PEAL) [11]. In this experiment, we choose 202 individuals (101 males and 101 females), each individual with 6 varying expressions. For each individual, we randomly select 4 of the face images for training and the rest 2 for testing. This database is substantially more challenging than the AR face database, since the number of subjects is now 202 but the train images is reduced to four per subject. We perform the similar experiments as on the AR face database.

From Tables 2 and 3, we can found that the registration does have a great influence on the recognition rates. Among the three methods, accurately registered images achieve the highest recognition rates; the performance of TPP is closing to that of accurately registered images; the accuracy of LP is much lower than TPP due to the insufficient number of

Table 4 Running time comparation between LP and TPP.

AR Face Database			CAS-PEAL Face Database		
Feature Length	LP (s)	TPP (s)	Feature Length	LP (s)	TPP (s)
30	22.3	2.7	30	17.6	3.3
130	46.5	4.9	56	21.8	4.5
300	62.0	6.4	168	53.4	5.3
560	82.0	8.1	700	132.7	7.4

training set.

Running time comparison between LP and TPP is shown in Table 4. Both the methods are tested under the same environment (2.8 GHz Mac Pro, 2 GB RAM, in Matlab). It can be seen from Table 4 that TPP is much faster than LP.

5. Conclusion

In this paper, we analyzed the limitations of LP, and proposed another registration method — Three-point positioning method. In order to demonstrate the ability of Three-point positioning method, we did experiments on public face databases and compared it with LP method. Experiments showed that, Three-point positioning method achieved higher recognition rates than LP. Besides, it is much faster than LP. As a result, TPP is a better registration method in practice.

Acknowledgments

References

[1] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.2, pp.210–227, 2009.
[2] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag., pp.21–30, March 2008.
[3] A. Majumdar and R.K. Ward, "Classification via group sparsity promoting regularization," ICASSP 2009, April 2009.
[4] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," ECCV, pp.448–461, 2010.
[5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," CVPR, pp.597–604, June 2009.
[6] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.5, pp.684–698, May 2005.
[7] A. Martinez and R. Benavente, "The AR face database," CVC Tech. Report, no.24, 1998.
[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," FGR, 2008.
[9] J. Zhou, Face recognition research in monitoring system, Thesis of Master's Degree, Tsinghua University, 2008.
[10] H. Kruppa, M. Santana, and B. Schiele, "Fast and robust face finding via local context," VS-PETS Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp.157–164, 2003.
[11] G. Wen, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," IEEE Trans. Syst. Man Cybern., vol.38, no.1, pp.149–161, 2008.
[12] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models — Their training and application," Comput. Vis. Image Understand., vol.61, no.1, pp.38–59, 1995.
[13] T.F. Cootes, C.J. Taylor, and A. Lanitis, "Multi-resolution search with active shape models," Pattern Recognit., vol.1, pp.610–612, 1994.
[14] J. Wang, G. Su, J. Liu, and X. Ren, "Facial feature points localization combining ASM and AAM," Journal of Optoelectronics Laser, (to appear).