From Rank-N to Rank-1 Face Recognition Based on Motion Similarity

Hui Fang
h.fang@mmu.ac.uk
Nicholas Costen
n.costen@mmu.ac.uk

Manchester Metropolitan University
Department of Computing and Mathematics,
John Dalton Building,
Chester Street, Manchester UK

Automatic face recognition systems have been investigated for many years as they can play important roles in the security of private information, access to computers, mobile phones, or to confidential location. To improve recognition performance, extra forms of information, such as depth or temporal cues, have been investigated by building 3D or dynamic models. Psychologists have proved that plastically deforming moving faces are significantly easier for humans to recognize than comparable static images; importantly, this is not simply an effect of additional samples [4]. Video based dynamic analysis is expected to contribute to form a counterpart of the psychological results. However, both psychological and preliminary computational studies show that this information is not strong enough to use on it's own. Thus we adopt a sequential approach to merge static and moving parameters.

Algorithm Design

The General Group-wise Registration (GGR) algorithm [3], is used to find facial correspondences. This is a iterative estimation method, updating the statistical model representing the image set and registering all the images according the new estimated model. A cost function is applied, including both shape and texture parts. Examples of the registration results are shown in Figure 1.

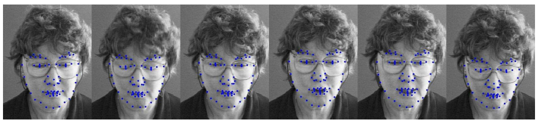

Figure 1: Representative frames from a BANCA individual. The left most image is the first frame of the sequence. The images have the landmark positions indicated on them.

A combined shape and texture model [2] is used to encode and represent the facial instances in the video sequences. Redundancies are removed using separate PCAs upon the shape and grey-level samples q_s and q_t, and the shape and texture parameters are combined to form a single vector for each image on which second PCA is performed. This gives a single feature vector x to represent each face instance,

$$x = {\Phi_c}^T \begin{bmatrix} W_s {\Phi_s}^T (q_s - \bar{q}_s) \\ {\Phi_t}^T (q_t - \bar{q}_t) \end{bmatrix}. \tag{1}$$

Static information is assessed by selecting nearest neighbours using the normalized correlation between parameters,

$$S_a = \frac{\bar{x}_i}{|\bar{x}_i|} \cdot \frac{\bar{x}_j}{|\bar{x}_j|}. \tag{2}$$

where $\bar{x}_i$ represents the average parameters of the probe and $\bar{x}_j$ represents the average parameters of one gallery sequence.

The motion analyzed here is subtle non-rigid distinctive facial changed. The variation is revealed when the images in one sequence are projected into the combined shape and texture subspace. The characteristic motion patterns are captured by using PCA performed upon all of the encodings x of a single sequence. The similarity of pairs of spaces is compared,

$$S_m = \max(\mathrm{trace}(\mathrm{colperm}(\Phi_i \cdot {\Phi_j}^T))); \tag{3}$$

this will involve calculating different permutations for a single probe sequence when comparing it with multiple gallery members. In addition, Φ_i and Φ_j may be truncated to remove noisy, low-variance eigenvectors.

The gallery size for S_m is based upon S_a, using the distribution of S_a as a function of the rank-position of the true target. The similarity of the probe and gallery declines rapidly and stabilizes after the true target is reached. The gallery is thus selected via the rank-n to rank-n+1 ratio.

Experiments

In Figure 2, the performance of S_m is compared with an auto-regressive model [1] and a mututal subspace method comparision [5] when two candidates, one of which is the true identity and the other the next nearest neighbour based on S_a, are involved in the selection process. Above about 20 dimensions, the S_m method is preferable, reaching 96.7% with two candidates.

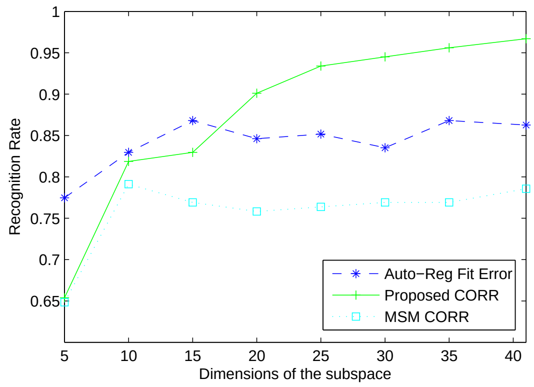

Figure 2: Recognition rate with 2 candidates using S_m alone.

To combine S_a and S_m, the ranked S_a ratios are used. As shown in Table 1, the flexible candidate selection can achieve 90.11 % accuracy, which is higher than using any fixed number of candidates. Note that S_a depressed by not using an identity-only space, so allowing S_m scope to improve performance.

	S_a only	2 Cand	3 Cand	Flex Cand
Rec. Rate(%)	85.16	87.59	88.23	90.11

Table 1: Recognition rates based on S_m with different number of candidates selected through S_a based recognition.

Conclusions

In this work, we proposed a sequential framework to integrate the motion feature in order to improve the facial recognition from the video sequences. Although the non-rigid facial motion is not fully predictable, it is still an important cue to help identification. With the help of permutated eigen-motion similarity as a subsidiary, the recognition rate is significantly improved based on a combined shape and texture appearance model. In future work we will explore fast feature finding and tracking techniques to speed up the process of modeling the face sequences.

[1] N. Campbell, C. Dalton, D. Gibson, Oziem D., and Thomas B. Practical generation of video textures the auto-regressive process. *Image and Vision Computing*, 22:819–827, 2004.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE T-PAMI*, 23(6):681–685, 2001.

[3] T. F. Cootes, C. J. Twining, V. Petrovic, R. Schestowitz, and C. J. Taylor. Groupwise construction of appearance model using piece-wise affine deformations. In *BMVC*, pages 879–888, 2005.

[4] K. Lander and L. Chuang. Why are moving faces easier to recognize? *Visual Cognition*, 23(3):429–442, 2005.

[5] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *F & G*, pages 318–323, 1998.