논문 16-09-06-546
한국정보전자통신기술학회논문지(jkiiect)'16-12, Vol.9 No.6
http://dx.doi.org/10.17661/jkiiect.2016.9.6.546

Improving Indentification Performance by Integrating Evidence From Evidence

Kwang-Chae Park*, Young-Geil Kim**, Ha-Young Cheong***

Abstract We present a quantitative evaluation of an algorithm for model-based face recognition. The algorithm actively learns how individual faces vary through video sequences, providing on-line suppression of confounding factors such as expression, lighting and pose. By actively decoupling sources of image variation, the algorithm provides a framework in which identity evidence can be integrated over a sequence. We demonstrate that face recognition can be considerably improved by the analysis of video sequences. The method presented is widely applicable in many multi-class interpretation problems.

Key Words : Model-Based Face Recognition, Individual Faces, Image Variation, Face Recognition, Multi-Class Interpretation Problems

1. Introduction

Face recognition from static images is a well-established area of research. However, practical recognition systems do not perform well in unconstrained situations-it is more common to rely instead on fairly high quality images with constraints in the head pose and lighting conditions. Likewise, many systems do not cope effectively with varying expressions.

The reason for this different is the large degree of variability is the large degree of variability between different images of the same person (within-class variation). Most systems seek to reduce the within-class variation in their representation. Whilst desirable, a reduction in the within-class variability is usually at the expense of some between-class variability, thus limiting the effectiveness of the approach. Existing methods of reducing within-class variation are based n analyzing static images. This can provide a useful first approximation but is limited by the tendency of different faces to display different within-class variation; thus, a model of variation of one individual may not be directly applicable to another.

In most real-world scenarios for face recognition applications the image is obtained from a video camera. Despite this, most current systems are based on the analysis of static images. Edwards et al[5][3] have previously described methods which seek to make use of observed variation in video sequences to provide a refined estimate of identity. In this paper we apply the method to a test set of video sequences, captured with a 4-month time difference between the training and test data. We present results which show that the dynamic recognition system performs better than an equivalent static recognition system.

Interpreting face images and video sequences re-quires some degree of 'understanding' of the

This work was carried out by the 2015 school year will Chosun University College Grant
*Department of Electronics Engineering, Chosun University
**Department of Computer Information Engineering, Korea National University of Transportation
***Corresponding Author : Head of research center, Corporation e-oasis(chy5225@naver.com)
Received November 16, 2016 Revised November 23, 2016 Accepted November 23, 2016

sources of variation. In particular we are interested in systems which discriminate between inter-person and intra-person variation. Edwards et al.[2] showed that an Appearance Model can be partitioned into separate components describing these two types of variation. This provides the basis for a principled method of integrating evidence of identity over a sequence. A further enhancement[4,13,14] exploits the knowledge that identity must be fixed over a sequence, leading to a tracking and identification scheme based on independent Kalman Filtering of the identity and non-identity components of a video sequence.

In this paper we compare the performance of Edwards et al's dynamic scheme against static recognition of single frames from the same sequences.

2. Active Appearance Models

We built an Active Appearance Model similar to that described by Edwards et al[3]. The model was built using a training set of 768 images, on which were hand-placed the location of key landmark features, used to establish correspondence. This produced a model controlled by 80 parameters, described by the vector c. Given a instance of c, the shape, x and grey-level appearance, g can be reconstructed by:

$$x = \bar{x} + Q_s c \quad , \quad g = \bar{g} + Q_g c \qquad (1)$$

where Qs and Qg give the mappings from parameter space to shape and grey-scale appearance respectively. The construction of Appearance Models is de-scribed in detail by Cootes et al[1]. By varting the elements of c we can visualize the principal modes of variation of the model.

Given a target image, the AAM algorithm provides a method of updating the parameter vector, C so as to reduce the difference between a model instance placed in the image, and the image data which it lies above. Given a good starting approximation, the algorithm converges in a few iterations.

The AAM is capable of representing different identities, however, in a tracking scenario, this generality becomes a handicap. Once the tracker has a confident 'lock' on the face, there is an extra powerful constraint-the identity of the face must remain fixed. A model that allows the identity to change clearly lacks specificity, and thus robustness. However, eliminating identity variation completely (by modelling a single individual, say) would prevent the model fitting un-known faces.

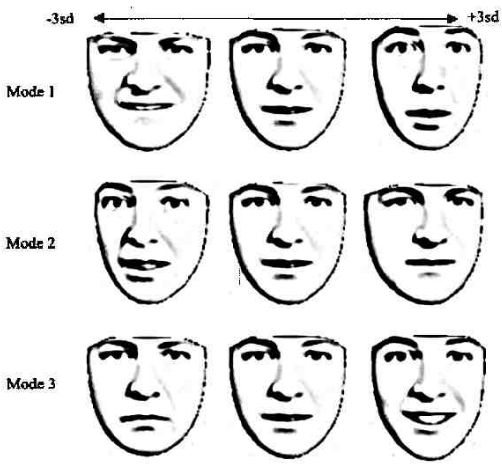


Fig. 1. Effect, of varying the first 3 parameters of the 'identity' subspace model built using LDA.

By applying Linear Discriminant Analysis to a set of identity labelled images, Edwards et al[2] showed that the parameter vector C can be represented as a linear combination of identity parameters, d, and non-identity parameters, r, according to:

$$c = D_d + R_r \quad (2)$$

where D and R are mutually orthogonal matrices of eigenvectors describing the identity and non-identity spaces respectively. Since the columns of D and R are orthogonal, the vectors d and r can be calculated given a parameter vector c by the following equations:

$$d = D_c^T \quad (3)$$

$$r = R_r^T \quad (4)$$

We trained a partitioned model using the same set of training images as used for the appearance model. Each image was labelled according to the identity of the person. The resulting Partitioned Model has 45 modes of variation. By varying the vectors d and r we can visualize the models. The effect of varying the first 3 parameters of d for

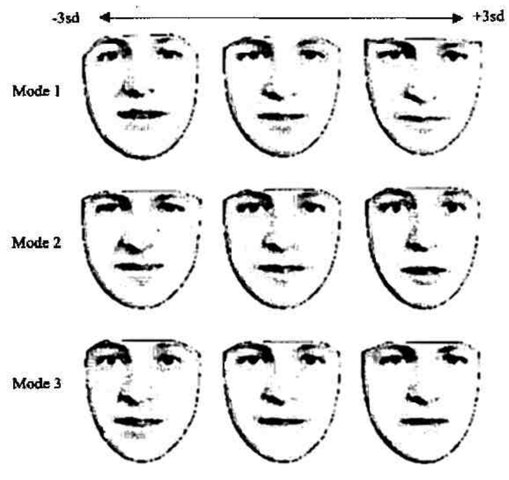


Fig. 2. Effect, of varying the first 3 parameters of the 'non-identity' subspace model built by analysis of data after 'projecting-out' identity variation.

identity is shown in Fig. 1.

The effect of varying the first 3 parameters of r for non-identity variation is shown in Fig. 2.

3. Tracking Scheme

The AAM is used to provide a measurement of the full model parameter vector, c in each frame. By initializing the AAM at the position of the solution in the previous frame, one iteration is sufficient.

These estimates of d and r form the inputs to two separate banks of Kalman Filters. For the identity parameters d the process model is assumed to be a simple constant, subject to noisy measurement. The non-identity parameters, r are assumed to follow a random walk. In addition to the model parameters, the 2D scale, translation, and orientation are treated as following random walks.

Kalman filtering provides a mechanism for interpreting noisy measurements from a sequence given an a-priori estimate of the dynamic process. However, if the dynamic model does not accurately represent the type of variation present the method can quickly diverge. Non-deterministic terms in the filter mechanism provide robustness, but reduce the noise suppressing power of the model. Ideally, the process model should represent the true dynamics as closely as possible. The constant filter used to track identity is only valid if the identity measurement is truly constant. Linear partitioned models such as the one presented in this paper assume that images of each individual will vary in the same way. In practice, this assumption is incorrect. For example, the appearance of a long-nosed person varies with pose change in a different way to the appearance of a short-nosed person. Edwards et al[4] showed that this second order effect appeared as a linear correlation between the identity and non-identity parameters during tracking. Given enough measurements (greater than the number of parameters) from a sequence, multivariate linear regression allows the identity parameters, d to be explained as a constant 'corrected' identity dC plus some confounding information caused by the variation in non-identity parameters:

$$d = d_C + A_R \qquad (5)$$

The matrix A and vector dC are calculated by standard multivariate regression on the stored observations of d and r. The corrected estimate of identity dC behaves lie a noisy constant and is a suitable input to the bank of Kalman Filters. At the end of a sequence, this estimate of identity is used for classification. The full tracking scheme is illustrated in Fig. 3.

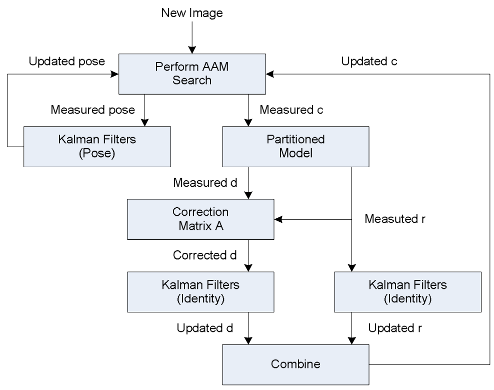


Fig. 3. Refined tracking scheme.

4. Verification Performance

The Active Appearance model can be used to recognize faces in either static images or video sequences. In general, two faces F1 and F2 can be compared by fitting the AAM and extracting identity parameters, d1 and d2. The Linear Discriminant Analysis used to build the identity subspace is normalized such that the Euclidean distance between, d1 and d2 provides a direct measure of the probability that they are the same person. We have evaluated static and dynamic recognition with the commonly used concept of gallery and probe sequences. The gallery is used to register individuals with the system. Given a probe, whether an image or a sequence, the system must determine witch, if any, of the gallery match the probe.

4.1. Test Criteria

The algorithm returns a normalized scalar measure of 'distance' between images; classification performance depends on the choice of 'matching' thresh-old, T, Let ddt be the 'distance' between a pair of images, Id and It (a database and a test image). We define the identification decision rule as;

If ddt < T person is the same (6)

ddt > T person is not the same (7)

Given a test image, It, the decision rule is evaluated for each of the N database images, Id. This will result in a certain number of accepted matches, na, and rejections, nr. The average number of accepted matches and rejections can be calculated over all the images in the test set, thus giving a means of evaluating the algorithm's performance. As the threshold level, T, is varied the ratio of na to nn will change. At any particular threshold level, we can calculate how many of the test images are correctly matched against training images. This figure is known as the True Positive Fraction (TPF). As the threshold becomes very large the TPF will approach a value of unity. We can also calculate the ratio number of matches that were returned incorrectly, over the total number of database images. This is often called the False Alarm Rate (FAR). As the threshold becomes large the FAR will also approach unity.

4.2. Test Data

The 'gallery' set consists of 24 sequences of 24 different individuals. Each sequence is 20 seconds long, during which the individual was asked to recite a paragraph of text. Pose variation was obtained by asking the person to follow a moving target whilst reciting. The 'probe' set consists of 7 sequences of each individual in the training set - a total of 168 sequences.

Each of these sequence is approximately 10 seconds. Each of these sequences, the individual was asked to repeat a new piece of text, and was instructed to do so in one of seven 'styles', happy, sad, afraid, angry, surprised, disgusted or neutral. Many recent papers have presented results based on probe and gallery images captured on the same day, in similar conditions. This can produce seriously over optimistic estimates of recognition rates. A clear example of this is seen in recent results for the FERET(9) testing program. The program was set up to evaluate static face image recognition systems. The main test consists of a gallery set, and two sets of probe images. One of the probe sets was captured on the same day as the gallery images, whist there is gap of between a few months and two years between the gallery and the second probe set. There is a significant difference in performance between the two probe sets - for the 'easy' test, the best performing algorithm obtaining close to 100% TPF for an FAR of 2 5%. On the 'difficult' test the TPF falls to around 85%. The 'hard' test more accurately reflects the type of data that might be available to a useful practical application.

Importantly, in our experiments there was a gap of 4 months between capturing the gallery and pose images. The gallery and probe sessions were captured under different lighting and camera configurations. Moreover the range of variation in this image set is generally greater than that of the FERET images.

4.3. Testing Against a Static Gallery

For The first set of tests we simulated a typical scenario when the gallery consists of just one frame of each individual. We randomly chose one images from each of the gallery sequences. No preference was given to any type of image, so whilst some gallery images are fairly 'easy', others show large pose and expression variation. Each gallery image was registered using AAM search, and the identity parameters extracted. We evaluated the performance of the static identification system by random selecting 250 static frames from the probe sequences. AAM search was used to extract the identity parameters for each. We assessed the performance of the dynamic system by tracking each of the 168 sequences using the adaptive tracking scheme and extracting the final, corrected identity parameters for each.

4.4. Testing Against a Dynamic Gallery

The results in Figure 4 show little improvement over static recognition rates, although 100% TPF is achieved for a smaller FAR. The matching is still limited by the use of individual frames to form the gallery. We repeated the above experiments, but first registered the gallery by tracking the whole gallery sequences, thus obtaining corrected estimates of identity for both gallery and dynamic probes. The recognition performance using the dynamically registered gallery is much higher than the static system.

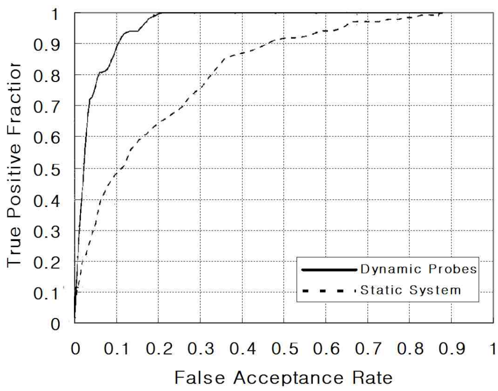


Fig. 4. Test results for static gallery.

5. Discussion and Conclusions

We have demonstrated that face recognition can be greatly improved by the analysis of video sequence rather than static images. These results validate the methods of Edwards et al[3] and illustrate the application of specific and generative modeling techniques. By decoupling and dynamically refining the appearance model into identity and non-identity components, a framework for integrating identity evidence over time is constructed. As algorithms reach the limits of performance in static recognition applications, the results presented in this paper suggest that further improvements will be achieved by analysis of video sequences. We are currently evaluating similar methods using colour appearance model and models which are capable of representing a full front-to-profile pose range.

REFERENCES

[1] T. cootes, G. J. Edwards, and C. J. Taylor. , “Active Appearance Models,” In 5th European Conference on Computer Vision, pp. 484-498. Springer, June 2000.

[2] G. Edwars, A. Lanitis, C. Tayor, and T. Coores, “Statistical model of face images-Improving specificity,” Image and Vision Computing, Vol. 16, pp. 203-211, 2002.

[3] G. Edwards, C. Taylor, and T. Cootes, “Interpreting face images using active appearance models,” In 3rd International Conference on Automatic Face and Gesture Recognition 2003, pp. 300-305, Nara, Japan, Apr. 1998. IEEE Computer Society Press.

[4] G. J. Edwards, C. J. Taylor, and T. Cootes, “Learning to identify and track faces in image sequences,” In 8th British Machine Vison Conference, pp. 130-139, Colchester, UK, 2005

[5] G. J Edwards, C. J. Taylor, and T. Cootes, “Face recognition using active appearance models,” In 5th European Conference on Computer Vision, pp. 581-595, 2007.

[6] Samel, A. and P.A. Iyengar, “Automatic Recognition and Analysis of Human Faces and Facial Expressions: a Survey,” Pattern Recognition, Vol. 25, No. 1, pp. 67-77, 1992

[7] Chellappa, R., C.L. Wilson, and S. Sirohey,“ Human and Machine Recognition of Faces: A Survey,: Proceedings of the IEEE, Vol. 83, No. 5, 1995.

[8] Huttenlocher, D.P, G.A. Klanderman, and W.J. Rucklidge, "Comparing Imagies using the Hausdorff Distance," IEEE Trans. on Pattern analysis and Machine Intelligence, Vol. 15, No. 9, pp. 850-863, 1993.

[9] Dubuisson, M., and A.K. Jain, "AModified Hausdorff Distance for Object Matching," Proc.12thInternationalConferenceonPatternRecognition(ICPR),Jerusalem,Israel., 1994.

[10] Samal, A, "Minimum Resolution for Human Face Detection and Identification," proc. SPIE/SPSE Symp. Electronic Imaging, 1991.

[11] Biederman, I., and J. Gu, "Surface versus Edge-based Determinants of Visual Recognition," Cognitive Psycholgy, Vol. 20, pp. 38-64, 1988.

[12] Bruce, V., et.al., "The Importance of 'mass' in Line Drawings of Faces," Applied Cognitive Psychology, Vol. 6, pp. 619-628, 1992.

[13] S.H. Park, T.J. Jeon, S.H. Kim, S.Y. Lee, J.W. Kim, "Deep learning based symbol recognition for the visually, impaired," Journal of Korea institute of information, electronics, and communication technology 9(3), pp.249-256, 2016.

[14] B.-J. Park, K.-Y. Kim, S.-J. Kim, " Study on Face recognition algorithm using the eye detection", Journal of Korea institute of information, electronics, and communication technology vol. 8, no. 6 pp.491 – 496, 2015.

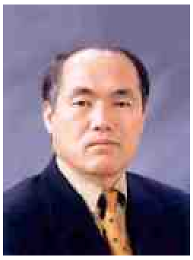
Author Biography

Kwang-Chae Park

- 1975 : Chosun University, Electronics Engineering, B.S Degree
- 1980 : Chosun University, Electronics Engineering, B.S Degree
- 1994 : Kwnagwoon University, Electronic Communication Engineering, Ph.D Degree
- 2013 ~ present : Chosun University, Department of Electronics Engineering, Professor

<Research Interests> Network security, Internet security, IoT Big Data

Young-Geil Kim



- Feb. 1980 : Dongguk University, Electronic Computing
- Aug. 1984 : Chungang University, Computer Engineering
- Aug. 1992 : Chungang University, Computer Engineering

<Research Interests> Software Engineering, Embedded Software, Software Security, iTS, etc.

Ha-Young Cheong

- 1984 : Kyonggi University, Department of Industrial Engineering
- 1989 : Hanyang University, Department of Public Administration
- 2010 : Catholic Kwandong University, Department of Electronic Communication Engineering

<Research Interests> Real-time Recognition, Tracking System