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

Abstract

In this paper, we present a sequential framework using facial motion information as a subsidiary to improve face recognition performance. As is generally known, reasonable static face recognition has been achieved based on subspace reduction techniques. In order to further improve performance, some extra cues, such as temporal variation, are investigated by building dynamic models. We propose a permuted similarity motion feature and integrate it into a sequential recognition system. This system can select the best candidate from the Rank-N candidates picked up in the recognition step based on static appearance parameters by using motion information. The recognition rate of the motion similarity is compared with the motion feature obtained from auto-regressive models to prove its efficiency. In addition, the sequential system achieves better performance when the motion information is integrated with the static appearance information in a flexible manner.

1 Introduction

Automatic face recognition systems have been investigated for many years because they have the potential to play important roles for the security of private information, access to computers, mobile phones, or to some confidential location. In recent years, researchers have explored more advanced methods to make these systems faster and more robust, so being suitable for real world applications.

A wide range of face recognition techniques, such as Principle Component Analysis [23], Linear Discriminant Analysis (LDA) [4], Elastic Bunch Graphic Matching (EBGM) [24], Active Appearance Models (AAM) [9], Independent Component Analysis (ICA) [3] and kernel-based techniques [20], have achieved reasonable results and each technique has its own advantages for identity decisions. Eigenfaces form the simplest algorithm to implement and a large number of techniques have been developed based on it; LDA proposes a linear projection with the maximum discriminant power; EBGM uses a collection of local features determine identity; AAM allowed combined shape and texture variation to be measured; ICA uses higher moments of data to find features which are independent with each other; kernel-based methods assume the projection space is nonlinear and design kernels to further project the data into a higher dimensional linear space. Often these techniques can be combined to produce more complex solutions.

In order to further improve the recognition performance, some extra forms of information, such as depth or temporal cues, have been investigated by building 3D models or dynamic models. Building a 3D model [6, 21] is highly dependent on the 3D acquisition equipment, while the hardware for this kind of machine limits the application of the model because the capture of the depth information is sensitive to external environmental conditions. In contrast, temporal information is much easier to obtain in that video cameras have become ubiquitous. Also, psychologists have proved that plastically deforming moving faces are significantly easier to recognize by humans than comparable static images; importantly, this is not simply an effect of additional samples [18, 22].

Video based dynamic analysis is expected to contribute to form a counterpart of the psychological results. Several frameworks have been designed to use this kind of motion information [1, 19, 26]. In this paper, we propose a sequential scheme to exploit facial motion to improve the recognition rate in a relatively large-gallery situation.

The rest of this paper is organized as follows. In section 2, we review several techniques to explore motion features. Section 3 introduces our framework and explains the proposed algorithm in detail. Experimental setup and results are shown in section 4. The conclusions are drawn in the last section.

2   Relevant Work

In recent years, dynamic information has been intensively investigated in order to predict human behaviors or improve face recognition systems [1, 5, 7, 12, 19]. A number of studies assume that the facial moving patterns of individuals are predictive [5, 7] so that regression based models can be built to capture the motion pattern. Conversely, others assume that the motion is un-predictive even though some familiar behavioral patterns can be learned to represent the variations an individual makes. Our proposed algorithm is based on the second assumption that the distinctive types of motion of the individuals are assumed to make contributions to identity.

A number of studies explore the use of predictive systems such as Auto-Regressive (ARM) and Hidden Markov Models (HMM). The auto-regressive model [7, 11, 12] is a typical dynamic system for learning the moving patterns of individuals. The state vectors α_t in the current frame can be approximated using a linear combination of several previous steps,

$$\alpha_t = \mathbf{w} + \mathbf{A}_1 \alpha_{t-1} + ... + \mathbf{A}_k \alpha_{t-k} + \mathbf{n} \tag{1}$$

where $\mathbf{w}$ represents the mean, $\mathbf{A}_i$ denote the transition matrices and $\mathbf{n}$ is zero-mean Gaussian noise. A least squares method can be used to estimate all the parameters to represent the behaviours. Alternatively, Liu and Chen [19] proposed training a set of HMMs, one per person, to describe dynamic facial information. The likelihood score of the sequence given each HMM is compared to allow recognition of the subject. Although the transitions between these states represent the motion patterns of the subjects to some extent, it was quite difficult to define the discrete states and quantize the parameters of the images.

Other studies have taken a non-generative approach. Edwards et al.[13] proposed an on-line framework to learn a linear relationship between the static facial appearance and motion information and so adjust the central appearance based on this relationship. This process improved recognition performance compared with simply taking the sequence-mean. The length of the moving sequence is an important factor when learning a robust relationship between the movement and appearance. It addition, the system does not actually model the

facial behaviour; discrimination is soley based on the adjusted static appearance. Arandjelović and Cipolla [1] report a system similar to that proposed here, although using an eigenface encoding. Relatively long sequences, where the predominate form of variation involves facial orientation, are encoded and their distributions are compared with regard to both their mean and variation. Notably high recognition performance was observed, but this depended upon the use of a kernel PCA to effectively encode the faces and also a procedure which compensates for different proportions of facial orientations in the sequences by physical modeling these changes. It is unlikely that this latter approach could be applied to sequences where the major variation is plastic facial changes (typically expressive or speech deformations).

In this paper, we propose a distribution based motion feature and integrate it into a sequential framework to identification. Based on the experimental results, we show that the non-rigid distinctive motion of individuals is useful for improving face recognition and the distribution based motion feature is more suitable than the predictive model based feature.

3 Algorithm Design

3.1 Framework Overview and Subspace Representation

3.1.1 Proposed framework

Common face recognition systems have a standard procedure to identify the subjects. When the face instance is located from an image, it is aligned with the faces in the gallery by using a rigid or non-rigid transformation. After re-sampling from the correspondent positions, the instance is projected into a subspace to find the nearest neighbor comparing with all the faces in the gallery.

A sequential framework is proposed to integrate the motion information into the recognition system. This structure is more suitable than a parallel structure because of the characteristics of the motion features. Although psychologists have proven that identification based on pure facial motion is possible, this is not strong enough to find the true subject correctly from a large number of distracters [16]. However, preliminary experimentation has shown that the motion features are quite useful for distinguishing subjects from a small number of candidates or for use in a verification framework [14]. Thus the subject can be recognized from small group of candidates when the system is confused if only the static identity features are used.

3.1.2 Salient points location and encoding model

The first and the most important step of encoding and analyzing facial movement is to solve the correspondence problem between the frames of a video sequence, and also across sequences. In this work, a registration framework is used to find the salient correspondences. AAM is a popular algorithm to locate the facial feature points, but the motion investigated here involves subtle non-rigid deformations which are hard to synthesize using a linear combination of a limited number of basis functions, as in the AAM. A number of algorithms use optical flow techniques to search for correspondences [8, 15] . However, ill-posed assumptions make it hard to track faces which suffer from the problems of occlusion and drift. The registration framework used here is more robust and accurate at locating the salient points than such other techniques.

The General Group-wise Registration (GGR) algorithm, extended from [10], is applied in this paper. This is a iterative estimation method to update the statistical model representing the image set and register all the images according the new estimated model. Thus this method breaks the chicken-and-egg problem that the majority of modeling and localization algorithms.

Figure 1: Representative stages of the iterated updating of the mean face model. Later stages are to the right correspondences location; note the increased level of detail.

The model used here is a simple mean shape and texture built by warping all the faces to the mean shape using a triangular Delauney mesh. Figure 1 shows how the registration improves and the model represents the face structure more accurately as different subjects are progressively aligned together. The matching step deforms each image by minimizing a cost function. A coarse-to-fine scheme is applied to increase the number of control points and optimize their position. In the final iterations, the points are moved individually to minimize the cost. The cost function includes both shape and texture parts,

$$E = \lambda \sum_i \left(c - \frac{0.5\|d_i - (\Delta d_i + d_{neig})\|}{\sigma_s^2} \right) - \frac{|r|}{\sigma_r} \tag{2}$$

where r is the residue between the model and the current image after deformation, σ_r and σ_s are the standard deviations of the residue and shape, c is a constant, d_i is the position of the ith control point, d_{neig} is the average position of the neighborhood around point i and Δd_i represents the offset of the point from the average mean shape. Examples of the registration results are shown in Figure 2.
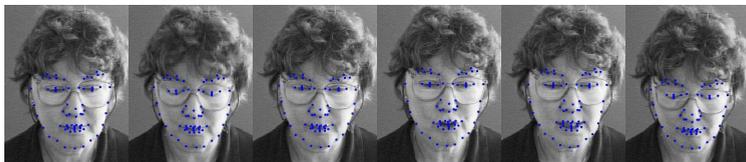

Figure 2: Representative frames from a BANCA individual. The left most image is the first frame of the sequence. The images have the landmark positions indicated on them

When the facial feature points are located, a combined shape and texture model [9] is used to encode and represent the facial instances in the video sequences. It is assumed that rigid facial movements contain relatively little distinctive information; as a result, it can be compensated for by a similarity transformation. The model then allows accurate

coding and recognition of examples. With the piece-wise warping procedure to an average shape, it ensures that the image-wise and face-wise coordinates of images are equivalent. Redundancies are removed by performing separate PCAs upon the shape and grey-level samples $\mathbf{q_s}$ and $\mathbf{q_t}$, providing shape and texture parameters. These are combined to form a single vector for each image on which second PCA is performed. This gives a single feature vector $\mathbf{x}$ to represent each face instance. Thus,

$$\mathbf{x} = \Phi_c^T \left[\begin{array}{c} \mathbf{W_s}\Phi_s^T(\mathbf{q_s} - \bar{\mathbf{q}}_\mathbf{s}) \\ \Phi_t^T(\mathbf{q_t} - \bar{\mathbf{q}}_\mathbf{t}) \end{array} \right] \tag{3}$$

where Φ_s denotes the eigenvectors on shape, Φ_t denotes the eigenvectors on texture, Φ_c denotes the eigenvectors on the combined features and W_s is a weight applied to the shape parameters to correct for differential variance on the two modalities.

3.2 Algorithm Details

3.2.1 Average feature based recognition

When a number of faces in the sequence are encoded on the manifold, the nearest neighbours are selected based on similarity measured by calculating the normalized correlation between the mean parameters of the probe sequence and the corresponding parameters of the gallery sequences,

$$S_a = \frac{\bar{\mathbf{x}}_i}{|\bar{\mathbf{x}}_i|} \cdot \frac{\bar{\mathbf{x}}_j}{|\bar{\mathbf{x}}_j|}. \tag{4}$$

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

The rank-n recognition rate based on these static average facial features extracted moving sequences and also that obtained from randomly selected individual frames is shown in Figure 3. This shows recognition for a total of 182 probes in a 66-member gallery. As suggested elsewhere [17], average parameters can achieve better results than a face vector derived from a single frame. Therefore, the average parameters are used to calculate the static similarity.

As expected, the cumulative match rate shown in Figure 3 increases markedly when several neighbours are included as being correct. If the motion information is also distinctive and stable across individuals, it may contribute to recognizing the correct target when the identity feature alone cannot correctly discriminate between individuals.

3.2.2 Motion feature based recognition

The motion analyzed here is subtle non-rigid distinctive facial motion. It is assumed that this kind of motion is a decisive cue for human identity. In order to ensure that the motion distribution as approximately Gaussian, the combined shape and texture appearance model is used to handle the non-linearity problem mentioned elsewhere [7]. This subspace can capture the data variation as a unimodal distribution shown in Figure 4 (this shows the encodings on the three highest-variance eigenvectors for two sequences of a single individual). Although subtle dynamic motion is a relatively weaker cue than the static facial features, it is important to confirm the identity if the facial appearance feature is similar for two different subjects.

When the moving images in one sequence are projected into the combined shape and texture subspace, the variation of these encoded representations reveals the movement of the
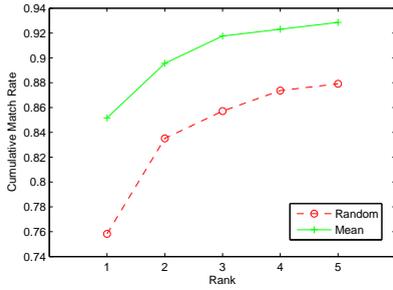
Figure 3: The rank-n recognition rate for static features. The average vectors obtained from sequences are more robust than the static single vectors.
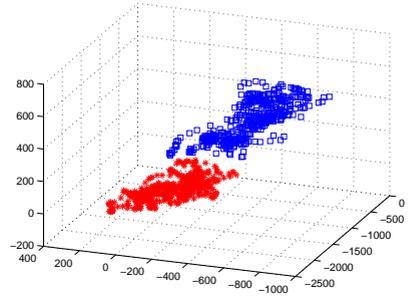

Figure 4: The subspace captures a uni-modal distribution of patterns and the distribution of two sequences of the same person look quite similar.

face; given the psychology it should be possible to perform recognition on this variation. The characteristic motion patterns are captured by using PCA performed upon all of the encodings $\mathbf{x}$ of a single sequence to yield a single set of eigenvectors Φ. Although there is no direct relationship between the motion patterns (a series of eigenvectors) and the specific behaviors of faces, some of the more important of variations actually represent some meaningful changes.

The similarity of pairs of spaces can be compared as

$$S_m = \max(\text{trace}(\text{colperm}(\Phi_i \cdot \Phi_j^T))), \tag{5}$$

where it should be noted that the ordering of the components extracted by the PCA will be determined by their associated eigenvalues and so the value of S_m is liable to be reduced by variation in the magnitude of different behavioral tendencies in different sequences of the same individual. This can be overcome by permuting the columns of the matrix describing the relationship between any two sequences before taking the trace so as to maximize S_m. It should be noted that 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 (identically) truncated to remove noisy, low-variance eigenvectors.

In Figure 7, the recognition rates of the proposed motion feature are shown based on pair of candidates, in each case there are the true target and the next most similar face, determined using S_a. As the variance in the appearance model increases, the distribution based motion more robustly identifies the subjects from a few candidates. As shown in Figure 8, the recognition rate drops when more candidates are involved in the candidates list would be expected.

3.2.3   Flexible candidate selection

Although promising performance can be achieved by a sequential framework when several candidates are selected based on S_a, it is important to decide how many candidates should be selected to achieve the best result.

Because S_a is more important for human recognition than S_m, it is plausible to decide the number of candidates for motion analysis based upon it. One possible method is to consider

the distribution of similarity measures as a function of the rank-position of the true target. As shown in Figure 5, the similarity of the probe and gallery (expressed as a ratio with the rank-1 gallery sequence similarity) declines rapidly and stabilizes after the true target is reached. Similarly, the true target and the distracters above it in the ranking have relatively constant similarity levels. These are the sequences which should be investigated as possible matches. It is possible to use the ratio of similarity between the rank-1 and rank-n targets to select a reduced gallery, as is shown in Figure 6. This records the distribution of ratios for the true targets and for the next nearest sequence, and, once converted to an ROC curve, suggests a criterion ratio of approximately 1.1. However, this is not necessarily appropriate, as the consequences of a false accept and false reject are not equal. A false reject (when the criterion is higher than the ratio for the target) ensures an incorrect overall match, while a false accept (where the criterion is lower than at least one distracter which is ranked after the target) only increases the size of the S_m gallery. As can be seen from Figure 8, S_m-only recognition declines as extra gallery members are added, reaching a final figure, with all faces included, of 87.36%. The criterion is thus decreased by 10%, increasing the number of targets included in the reduced gallery.
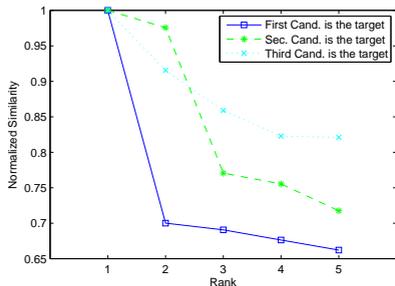

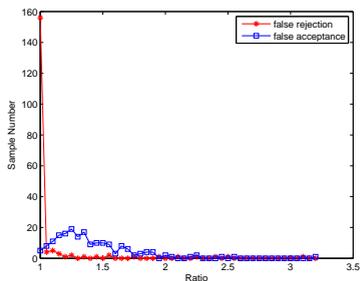Figure 5: The S_a ratio tendency with rank position, against the location of the true target.

Figure 6: The distribution of (inverted) S_a ratios for target and target +1 sequences.

4 Experiments

The proposed framework is evaluated based on two moving face databases, BANCA database [2] and Manchester University (MU) database. Both of these include frontal face sequences with slight rigid head motion. The BANCA database includes two groups, $g1$ and $g2$, each of which has 26 subjects. In addition, there are another 10 subjects who are only used for the ensemble. For each subject, 8 sequences, one for the gallery and 7 probe sequences, were collected. Each such sequence has approximately 500 frames. Because of the limited number of subject, extra data from the MU database were added, which includes 40 subjects, 4 sequences for each. The group $g1$ and an extra 10 BANCA subjects are used as an ensemble to build the appearance model and for calibration, with 5000 grey-level samples per face and 68 feature points. The subjects from the other group, $g2$, and the MU database, giving of total of 66 individuals, are used to form the gallery and probes for the identification experiments. The individual eigenmodels which provide the motion-pattern parameters are each derived from approximately 500 observations, each of which has 41 dimensions (98%

of the variance was encoded by the general appearance model).

4.1 Feature Comparison

4.1.1 Feature comparison with ARM

In Figure 7, the performance of S_m is compared with an auto-regressive model feature 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. The auto-regressive model provides an error based feature. The behavior of a specific individual is captured using a two-step ARM [7]. When the two candidates are selected using S_a, the encoding parameters of the probe sequence are fitted in the two specific models and the sum of the absolute difference is used to decide which model is better at predicting the probe sequence. It can be seen that the ARM feature is an efficient method of recognizing the face from two candidates. However, the recognition rate fluctuates around 85%, significantly beneath that of S_m. It appears that the individual facial behavior is only partially predictive; this may reflect the relative consistency of vocal behaviour in the sequences.

Although the individual facial behavior is not fully predictive, some distinctive patterns still exist to distinguish the faces. If the motion of the face can be described based on the eigen decomposition method, the relationship between the correlated motion patterns is expected to be high while that between uncorrelated motion patterns can be treated as random noise. Therefore, the feature should become more robust when more dimensions are used, which allows more correlations between motion patterns. As a result, it can be seen that the S_m recognition rate reaches 96.7 % with two candidates. Notes that the results are based on the calibration set, which has 26 subjects, 7 sequences for each.
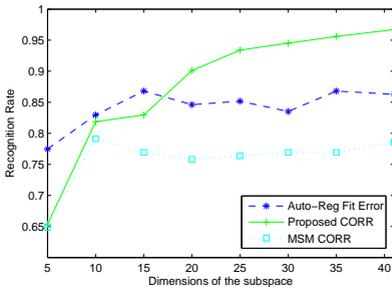

Figure 7: Recognition rate with 2 candidates using various motion models.
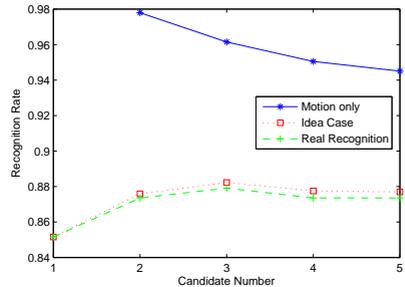
Figure 8: The S_m recognition rate with varying numbers of candidates.

4.1.2 Feature comparison with MSM

Another popular method of comparing pairs of distributions is the mutual subspace method (MSM) [25]. The first three principal patterns between the two PCA subspaces are compared to find the pairing which is most highly correlated. This method is extended based on the combined shape and texture model. A comparison of the proposed algorithm and this method is shown in Figure 7. It can be seen that the permuted correlation yeilds a notably higher discrimination with the subtly varying sequences of the type used here.

4.1.3 Comparison with the ideal case

In Figure 8 it is assumed that the true identity is always included in the candidate list. As the number of candidates increases, it can be seen that the recognition rate drops.

However, the stage using S_a cannot guarantee that the true client is always included in the candidate list. Therefore, the recognition also relies on S_a even though S_m achieves high recognition rates when selecting from a relatively small number of candidates. In Figure 8, the ideal recognition rate (calculated as the product of discrimination by S_a and S_m) on different number of candidates is shown by the red dotted line when S_a is selecting a variable number of candidates. The actual recognition rates based on successive selection by S_a and S_m (shown by the green line), and also given in Table 1, are in close accordance with the ideal results.

4.2 Candidate selection scheme

To balance ensuring that the true identity is present in the candidates list with avoiding unnecessary candidates, the candidate selection scheme introduced in previous section is used. If the recognition based on S_a is strong enough to determine the true client, it is unnecessary to pass several candidates to S_m. Conversely, more candidates should be selected and discriminated by S_m. As shown in Table 1, the flexible candidate selection can achieve 90.11 % accuracy which is higher than using any fixed number of candidates.

	ID Para. Only	2 Cand	3 Cand	4 Cand	5 Cand	Flex Cand
Rec. Rate(%)	85.16	87.59	88.23	87.75	87.7	90.11

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

4.3 Summary and discussion

Motion features are important cues to help people or computers on face recognition. We have designed a framework simulating phenomena concerning how humans recognize the face based on motion information. Because of the relatively weak nature of the motion cures, it is not possible to use them with large galleries. The proposed motion feature also has been proven to be efficient as a subsidiary in the sequential framework.

Furthermore, this framework can be easily extended to other subspace algorithms. Due to the small number of the subjects in most video based face databases, most more advanced subspace algorithms, such as an LDA space or SVM-type discrimination planes, suffer from ceiling effects. In another words, it is hard to prove the efficiency of motion feature because the correlation between the identity features has achieved a nearly perfect score. If a larger database can be generated, this framework would be tested based on those advanced techniques.

5 Conclusion and Future work

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. Further, we will test the system on a larger database and use advanced models to prove its efficiency.

6 Acknowledgements

References

[1] O. Arandjelović and R. Cipolla. An information-theoretic approach to face recognition from face motion manifolds. *Image and Vision Computing*, 24:639–647, 2006.

[2] E. Bailly-Bailliere, S. Bengio, F. Bimbot, and et al. The banca database and evaluation protocol. In *Proc. ICAVBPA*, pages 625–638, 2003.

[3] M. Bartlett, J. Movellan, and T. Sejnowski. Face recognition by independent component analysis. *IEEE Transations on Neural Networks*, 13(6):1450–1464, 2002.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.

[5] F. Bettinger, T. Cootes, and C. Taylor. Modelling facial behaviours. In *Proc. British Machine Vision Conference*, pages 797–806, 2002.

[6] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 25(9):1–12, 2003.

[7] 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.

[8] A. K. R. Chowdhury and R. Chellappa. Face reconstruction from monocular video using uncertainty analysis and a generic model. *Computer Vision and Image Understanding*, 91:188–213, 2003.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[10] 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 *Proc. British Machine Vision Conference*, pages 879–888, 2005.

[11] G. Doretto and S. Soatto. Dynamic shape and appearance models. *IEEE Transactions on Pattern Analysis and machine Inteligence*, 28(12):2006–2019, 2006.

[12] G. Doretto, A. Chiuso, Y. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, 2003.

[13] G. Edwards, C. Taylor, and T. Cootes. Improving identification performance by integrating evidence from sequences. In *Proc. Computer Vision and Pattern Recognition*, pages 1486–1491, 1999.

[14] H. Fang and N. P. Costen. Behavioural consistency extraction for facial verification. In *Cross-Modal Analysis*, pages 291–306, 2009.

[15] H. Fang, N. Costen, D. Cristinacce, and J. Darby. 3d facial geometry recovery via group-wise optical flow. In *Proc. Face and Gesture*, 2008.

[16] H. Hill and A. Johnston. Categorizing sex and identity from the biological motion of faces. *Current Biology*, 11:880–885, 2001.

[17] R. Jenkins and A. M. Burton. 100% accuracy in automatic face recognition. *Science*, 319(5862):435, 2008.

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

[19] X. Liu and T. Chen. Video-based face recognition using adaptive hidden markov models. In *Proc. Computer Vision and Pattern Recognition*, pages 26–33, 2003.

[20] J. Lu, K. Plataniotis, and A. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Transations on Neural Networks*, 14(1):117–126, 2003.

[21] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 28(1):31–43, 2006.

[22] A. O'Toole, D. Roark, and H. Abdi. Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6(6):261–266, 2002.

[23] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[24] L. Wiskott, Fellous J., N. Knueuger, and C. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 19 (7):775–779, 1997.

[25] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *Proc. Face and Gesture*, pages 318–323, 1998.

[26] S. Zhou, V. Krueger, and R. Chellappa. Probabilistic recognition of human faces from video. *Computer Vision and Image Understanding*, 91:214–245, 2003.