Selection of localised salient features from a single facial image

M.A. Grudin, P.J.G. Lisboa and D.M. Harvey

Abstract: A new application of dynamic link architectures is proposed for the automatic selection of key features in multiple instances of facial images, from manually tagged features in a single facial image of each person. This method is evaluated for its accuracy in face recognition, determined by the frequency with which the model graph fitted to a single instance of an individual's face achieves the closest match to other images of the same person. In addition to the potential of this method for face recognition, the methodology for feature alignment is a practical enabler to recognise emotion in affective computing. The system was first trained to evaluate typical intrapersonal variations of facial features on a training subset with ten facial images from six individuals in the Manchester Face Database (Lanitis *et al.*, 1993). After the training stage, facial features in an image of a new face were assigned the intrapersonal variations obtained for the corresponding features during the training stage. The saliency measure for each local image feature was then computed within a Bayesian framework and the accuracy of face recognition was evaluated with a further three images each from 24 people, taken from the same data set. A refinement of the saliency framework that used only a subset of local features for face recognition further increased the accuracy of face recognition on the same test database.

1 Introduction

For more than 20 years, face recognition has been an active area in computer vision research [1]. Over the years, researchers started to consider not only interpersonal variations between facial images from different subjects [2], but also intrapersonal variations [3]. Unlike interpersonal variations, which are computed between images of different people, intrapersonal variations are measured on several images of the same person. Examples of such work include Lanitis *et al.* [4], Krüger [5] and Moghaddam and Pentland [6]. The latter work used a maximum-likehood mechanism for face location with two types of density estimates, a multivariate Gaussian for unimodal distributions and a mixture-of-Gaussians model for multimodal distributions.

Consideration of interpersonal variations has been most successful when the variations were computed on the 'high-level' facial features rather than the 'low-level' local image features. Indeed, although both types of features can correspond to the same image region, it is the high-level facial features whose change in appearance can be predicted and estimated. However, using only a limited set of facial features is not sufficient to encode highly personal features like birthmarks and so on, which can appear anywhere on the face.

The majority of research has concentrated on using either the several high-level facial features or the low-level image features only. In this work, we aim to utilise both representation types, by implementing a locally matched wavelet pyramid of image filters. Further, the proposed method computes saliency of low-level image features using information from the corresponding high-level facial features. Image features with low saliency are discarded. The method is evaluated for the accuracy with which each model graph achieves the best fit for multiple images of the same subject.

A potential further application of the proposed methodology is to utilise the capability for automatic feature alignment as a tracking mechanism for emotional mapping in human–computer interaction through affective computing [7–9], for which adaptation of key parameters such as mouth location and mouth boundaries is required.

2 Overview of the approach

Despite the complexity of the face recognition task, human observers are considered to be experts in recognising faces. We can pick the most distinguishing facial features from a single image of a person and we appear to do this using a priori knowledge to direct their attention to the areas that are very different between faces of different people and yet preserve certain predictable properties between different appearances of the same face [10].

Here, it will be assumed that corresponding facial features of different faces often exhibit similar behaviour (a similar assumption has been made in Krüger [5]). On the basis of that assumption, we can use a small number of people to train the system to estimate intrapersonal variations of the high-level facial features. Once trained, the system assigns to facial features in an image of a new person, the previously learned values of intrapersonal variations.

IEE Proceedings online no. 20045197

doi:10.1049/ip-vis:20045197

Paper first received 30th September 2004 and in final revised form 12th September 2005

M.A. Grudin and D.M. Harvey are with the General Engineering Research Institute, Liverpool JMU, Byrom Street, Liverpool L3 3AF, UK

P.J.G. Lisboa is with Computing and Mathematical Sciences, Liverpool JMU, Byrom Street, Liverpool L3 3AF, UK

E-mail: d.m.harvey@livjm.ac.uk

However, even though some facial features are stable, they might not have sufficient interpersonal variations to discriminate between different people. Therefore both the estimated intrapersonal and interpersonal variations are needed to select a subset of the most discriminating image features. The proposed approach uses a Bayesian function to estimate the discrimination power of each image feature on the basis of the value of that feature and both the estimated intrapersonal and interpersonal variations in the appearance of that feature. The subset of features with the highest discrimination power is retained for future face recognition, whereas the other features are discarded.

Both the intrapersonal and interpersonal variations are computed for each feature on a training subset of faces. For computations of the interpersonal variations, the eye coordinates of the subjects are aligned and a single image per subject is used. The intrapersonal variation is computed on different images of the same person and averaged across the training set by mapping the distribution to a shape-free form, which is discussed in detail in Section 4.

Once those distributions are computed, they are used with each new face to select the most discriminative features. The interpersonal variation is used directly, whereas the shape-free distribution of intrapersonal variation is warped to correspond to the facial features of the new face. The most discriminative features are retained and used for subsequent face recognition.

To perform this experiment, a hierarchical graph-matching algorithm has been created. It differs from the previously described approaches [11]. This graph matching has some advantages over the previously described techniques and can be used on its own, but the further feature selection achieves still better performance. For the experiment, a separate model graph has been created for each person, with key facial features manually identified.

3 Matching of hierarchical attributed graphs

3.1 Hierarchical graph architecture and the face location stage

The dynamic link architecture (DLA) [3] is a powerful and flexible method for image analysis which addresses the problem of preserving topological information during shape deformations and changes in viewing position. The DLA constructs a regular 2D grid, whose nodes contain a multiresolution image description in terms of localised spatial frequencies. Each node has several feature detectors, based on modified Gabor-based wavelets [4], which describe the grey-level distribution locally with high precision and more globally with lower precision. The grid nodes are connected with elastic links, which can be distorted to account for changes in face shape. The final matching cost is a combination of the cost of matching each node in the grid together with a measure of the amount of grid deformation.

The principles of the original DLA have been used for the analysis of facial images, for instance, by Wiskott [13] and Krüger [5] who associate nodes of their attributed graphs with certain high-level facial features. Both methods investigate intrapersonal variations in appearance of facial features to improve the recognition performance. Würtz [11] analyses information on different image resolutions using a hierarchical graph structure, removing background graph nodes to achieve successful recognition independently of the background.

Visual information contains image data at multiple resolutions, and the discrimination confidence of facial features may depend on the resolution at which those features are observed. To analyse the statistical properties of image features on different scales, we use hierarchical attributed graphs, similar but distinctive from earlier research [11]. The adapted hierarchical graph structure contains six levels, with each level L containing $N_L \times N_L = 4^L$ nodes [13]. The sampling distance is set to $d_L = 2^{-L} \times p$, where p is a constant positive value.

In our approach, the graph structure contains both hierarchical and spatial links. Hierarchical links exist between each parent node and its children nodes. When the face is located and the position of the parent node is optimised, the child nodes are initially positioned with respect to their parent. This inheritance of coarse-scale information by fine-resolution grids reduces the risk of the child nodes being trapped in local minima of the cost function. As using only hierarchical links fails to preserve topological information [13], we also use spatial links between adjacent nodes in each level.

Each graph node contains amplitudes of the image responses to the complex Gabor-based wavelets at $D = 6$ orientations. The wavelet bandwidth is set to one octave and the spatial extent σ_L is set proportional to the sampling distance on that level, d_L [14]. The image responses are combined in a feature vector $\boldsymbol{k} = \{k(\alpha)|\alpha = 1, \ldots, D\}$, with the length of each feature vector normalised to $\|\boldsymbol{k}\| = 1$ in order to reduce illumination dependency.

For each person, a single model hierarchical graph is obtained from a forward-looking facial image without expressions. The graph is manually centred over the face with respect to the high-level features that are identified manually using landmarks (Fig. 1*a*). Altogether, 33 landmarks corresponding to major important facial features are assigned. This procedure is carried out only once for each individual in the database and for each new individual. Therefore for each new person in the database, only one image needs to be feature-marked manually, making the addition of new subjects very practical.

Information about locations of the facial features is used by the feature selection process, which is described in Section 4. The landmarks are also used to remove background nodes (Fig. 1*b*). Unlike in Würtz, [11], nodes over the hair region on top of the head are not removed; instead, because of their smaller discrimination power, they are expected to be automatically discarded by the feature selection mechanism described subsequently.

Graph matching is performed in a top-down fashion. Each level L in the pyramid is considered as a grid $M^L = \{\boldsymbol{k}^L_{ij}| \; i = 1, \ldots, N_L, \; j = 1, \ldots, D\}$, where i and j are the 2D grid coordinates. The corresponding subset of image points can be denoted as a subgraph

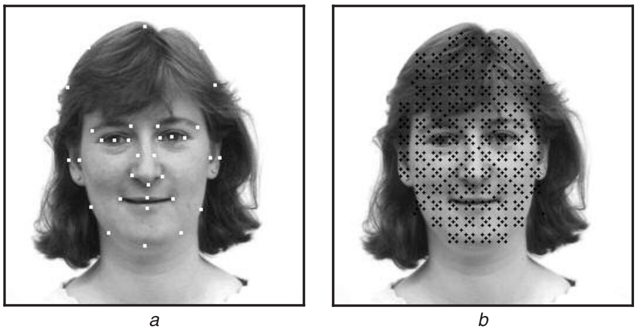

Fig. 1 *Single model hierarchical graph*

a Manual identification of facial features
b Image sampling

$I^L(\boldsymbol{x}) = \{\boldsymbol{i}^L_{ij}\}$, where $\boldsymbol{x}$ denotes the image coordinates of the top-left corner of subgraph I^L. The goal is to locate a set of coordinates $\boldsymbol{x}$ that minimises the cost S of matching M^L and $I^L(\boldsymbol{x})$.

For each level L of the hierarchical graph, the cost S of matching an extant grid M^L, to image-derived features for a grid centred at $\boldsymbol{x}$, $I^L(\boldsymbol{x})$, is computed as

$$S_{\text{grid}}\left(M^L, I^L(\boldsymbol{x})\right) = \frac{1}{N_l^2}\sum_{i,j=1}^{N_l} S\left(\boldsymbol{k}^L_{ij}, \boldsymbol{i}^L_{ij}\right) \tag{1}$$

The cost of matching each pair of nodes in M^L and $I^L(\boldsymbol{x})$ is defined as Euclidean distance

$$S\left(\boldsymbol{k}^L_{ij}, \boldsymbol{i}^L_{xy}\right) = \left\|\boldsymbol{k}^L_{ij} - \boldsymbol{i}^L_{xy}\right\| \tag{2}$$

We use the conventional procedure of matching attribute graphs [3], which consists of two stages. At the first stage, the face is located using coarse rigid grids, until the cost function S' is below a threshold T_L. At the second stage, described in the next section, the grid M^L is distorted to best match the image subgrid $I^L(\boldsymbol{x})$.

The goal of the first stage is to find the shift $\boldsymbol{x}$ of $I^L(\boldsymbol{x})$ within the image, which minimises the matching cost of the first stage $S'_{\text{grid}}(M^L, I^L(\boldsymbol{x}))$. If

$$S'_{\text{grid}}\left(M^L, I^L(\boldsymbol{x})\right) \le T_L \tag{3}$$

the face is considered as being located.

In practice, all faces in the training subset of the facial database were successfully located using only level $L = 2$. For that reason, levels $L = 0$ and 1 of hierarchical graphs were discarded. The value T_2 has been set to the maximum of all values $S'_{\text{grid}}(M^2, I^2(\boldsymbol{x}))$ for the faces in the training subset of the facial database and that value was consequently used to locate faces in the test subset.

3.2 Distortions of the elastic links in the hierarchical graphs

After the face is located, the second stage is performed by successively matching the remaining higher-resolution grids. At this stage, small deformations $\boldsymbol{y}$ of the links between the nodes are allowed in order to adjust the grid to changes in the underlying image pattern. The original elastic graph architecture [3] uses this distortion information as a part of the computation of the total cost of matching. In that approach, the spatial grid is designed so that the amount of facial distortion is typically less than the sampling distance of the grid (often a fraction of that distance). This property of the DLA ensures that the search for the best match is usually limited to a small image area.

This is no longer possible in the case of the hierarchical attributed graphs, because the amount of distortion, when measured in terms of the sampling distance, is inversely proportional to the grid resolution. Indeed, for high-resolution grids, the amount of distortion may be much larger than the sampling distance. This correspondingly increases the computational complexity of matching the high-resolution grids.

Another problem that must be considered when matching graphs on multiple resolutions is that the distortions are distributed quite unevenly over the facial surface. They are typically limited to several localised areas such as the mouth. The presence of large distortions makes it more difficult to use information about the grid distortions in the calculation of the matching cost.

As a solution to the increased complexity of the second stage of the hierarchical graph matching, we propose a new procedure of mapping refinement. Previously developed graph-matching methods use a random refinement order [3] or centrifugal refinement order [11]. In the described implementation, positions of nodes with higher matching correspondences are refined first. This order ensures that even in the presence of significant distortions the grid is anchored to the features that provide better correspondence.

In addition, links of the better-matching nodes are allowed to have larger distortions in order to maximise the search space for the links that are most likely to find the correct match. The nodes with higher matching cost are restricted to searching smaller areas: it is expected that most of the nodes will find their correspondence provided that the better-matching nodes are positioned correctly. This refinement procedure has shown good performance under a wide range of distortions, such as changes in face expression and head rotations (Fig. 2).

This matching procedure reduces the computational complexity of the second stage, although it does not deal with the fundamental problem of large distortions. We propose that future implementations of the elastic graph matching permit different distortions of the grid links depending on the facial region that corresponds to that part of the grid. Meanwhile, the 'unclassified' distortion information is not used in the calculation of the matching cost.

3.3 Computation of the total matching cost

During the second stage of the graph-matching process, the position of each node is optimised within the neighbourhood $a_L \times a_L$ around the original position of the node. The goal here is to match the graph nodes to a small neighbourhood of pixels in the image finding the most similar feature vector around the position $\boldsymbol{y}$ at the previous level, namely

$$S''\left(\boldsymbol{k}^L_{ij}, \boldsymbol{i}^L_{ij}\right) = \min_{y'}\left\{S'\left(\boldsymbol{k}^L_{ij}, \boldsymbol{i}^L_{ij}(\boldsymbol{x}+\boldsymbol{y})\right)\right\},$$
$$\text{where } |\boldsymbol{y}| < \frac{a_L}{2} \tag{4}$$

The values of a_L have previously been established for each level independently using the training set of images. The final cost of matching grid M^L is computed as

$$S''_{\text{grid}}\left(M^L, I^L(\boldsymbol{x}+\boldsymbol{y})\right) = \frac{1}{N_L^2}\sum_{i,j=1}^{N_l} S''\left(\boldsymbol{k}^L_{ij}, \boldsymbol{i}^L_{ij}\right) \tag{5}$$

The graph matching cost is computed as a weighted sum of the costs of matching the three finer-resolution levels. If the

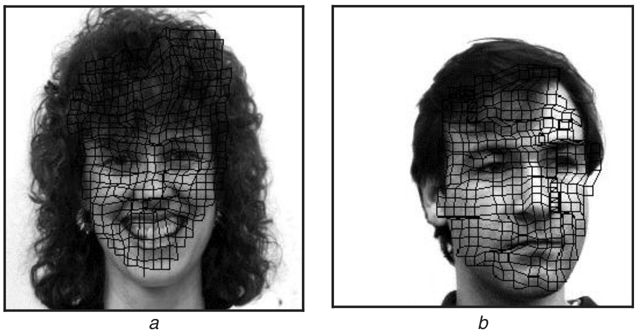

Fig. 2 *Distortions of high-resolution grids*

a Facial expression
b Head rotation

total cost of matching any single model graph is less than a given threshold T_{total}, then the graph is considered a successful fit to the facial image.

4 Incorporation of face-specific information

Implementation of the hierarchical attributed graphs, as described in the previous section, is a generic method that is effective for object recognition. However, modifications of the method to include class-specific information is expected to improve the recognition performance. In the case of faces, consideration of intrapersonal and interpersonal variation of facial features is used to estimate the discrimination confidence of different features. The less discriminative features are removed from the attributed graphs.

Consider how the posterior probability $P(c^m|k_{ij}^L(\alpha))$ of feature $k_{ij}^L(\alpha)$ indicating a presence of person c^m can be estimated using the Bayesian decision rule

$$P\Big(c^m|k_{ij}^L(\alpha)\Big) = \frac{P\Big(k_{ij}^L(\alpha)|c^m\Big)P(c^m)}{P\Big(k_{ij}^L(\alpha)\Big)} \qquad (6)$$

where $P(k_{ij}^L(\alpha))$ the probability of occurrence of feature $k_{ij}^L(\alpha)$, $P(k_{ij}^L(\alpha)|\ c^m)$ the probability of person c^m indicating the presence of feature $k_{ij}^L(\alpha)$ and $P(c^m) = 1/N$ the probability of person c^m being present $(m = 1, \ldots, N)$. Then the discrimination power $\Delta(c^m\ |\ k_{ij}^L(\alpha))$ of feature $k_{ij}^L(\alpha)$ to indicate the presence of person c^m is proportional to the maximum value of the posterior function within the limits of $\pm 1.96\sigma_{ij}^L(\alpha)$ from the mean value $\overline{k_{ij}^L}(\alpha)$, consistent with a 95% confidence interval in the case of normally distributed data

$$\Delta\Big(c^m|k_{ij}^L(\alpha)\Big) \sim \max\left[\frac{P\Big(k_{ij}^L(\alpha)|c^m\Big)/N}{P\Big(k_{ij}^L(\alpha)\Big)}\right]_{\overline{k_{ij}^L(\alpha)}\pm 1.96\sigma_{ij}^L(\alpha)} \qquad (7)$$

To estimate the values of $P(k_{ij}^L\ (\alpha))$ and $P(k_{ij}^L\ (\alpha)|\ c^m)$ for each $k_{ij}^L(\alpha)$, we need to know the interpersonal and intrapersonal distributions of facial feature appearances, correspondingly. We assume that, given a large number of samples, both distributions will have a Gaussian form. Then, those distributions can be reconstructed provided the mean of each distribution and its standard deviation are known.

The values of the mean $\overline{k_{ij}^L}(\alpha)$ and of the standard deviation $\varsigma_{ij}^L(\alpha)$ for the interpersonal distribution of each image feature $k_{ij}^L(\alpha)$ are obtained from all the model graphs in the training set (note that the interpersonal variation is computed for image features, whereas the intrapersonal variation is computed for facial features). The model graphs are aligned so that the point between the eyes in all graphs corresponds to the same physical location. Because after such alignment the graph nodes typically do not correspond to each other, the feature values between the graph nodes are interpolated on the basis of the values of adjacent nodes.

Unlike with the interpersonal distribution, the standard deviation of the intrapersonal distribution $\sigma_{ij}^L(\alpha)$ of facial features is computed separately for each face in the training set. For each person in that set, we obtain the intrapersonal distribution of the matching cost by matching the model graph to the other images of that person. A separate distribution is stored for each graph node. The standard deviation of this distribution indicates the degree of stability of the facial feature that corresponds to a particular node. The values of the intrapersonal distribution between the points that correspond to the graph nodes are interpolated on the basis of the values of the adjacent nodes.

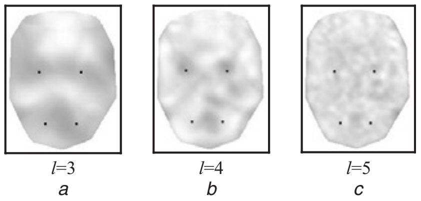

Fig. 3 *Shape-free distribution of intrapersonal variation for hierarchical levels l = 3, 4, 5*

The eyes and the mouth are denoted by dots. Large intrapersonal variations are illustrated by the lighter regions

However, this intrapersonal distribution is valid only for that person, because in the other person's model graph, the nodes will not necessarily correspond to the same facial features. To remove this dependency on the location of the facial features, or shape of the face, the intrapersonal distribution of each face in the training set is warped to a shape-free form. In practice this is done by warping the facial landmarks obtained from the model graph of a particular face to pre-defined (shape-free) locations of facial landmarks. The shape-free model used in this method is similar to those developed in Chellappa *et al.* [1] and Craw and Cameron [16], but the important distinction is that it contains the distribution of intrapersonal variations rather than purely the grey-scale distributions.

In the experiment, six intrapersonal distributions have been generated based on the six persons in the training set. Those distributions have been warped to a shape-free form. The resulting average distribution of such shape-free intrapersonal variations is illustrated in Fig. 3, with lighter regions corresponding to larger intrapersonal variations. The result confirms the assumption that some facial features exhibit significantly larger variations than the others do. The unexpected conclusion is that certain facial features appear stable on some scales, but vary significantly on different scales. On all scales, the hair region exhibits large variations. The nose also appears as less reliable, apparently because of significant differences in its appearance under side-to-side head rotations. Because the nose is the farthest feature from the spinal cord, which is the centre of such rotations, although it is the closest to the camera point in the face, its projections under different rotations are seen as containing more variations. The eyes and the mouth typically exhibit less intrapersonal variation in their appearance. However, their stability depends on the image resolution, for example, the iris and the lid movement on the high resolution contributes to higher intrapersonal variations of the eye region.

Once created, this average shape-free distribution is used to estimate intrapersonal variations of a new face. In practice, this is done by warping the shape-free distribution of intrapersonal variation 'back' to the new face, so that the pre-defined shape-free landmarks correspond to the facial landmarks in the face. With estimated intrapersonal variation $\sigma_{ij}^L(\alpha)$ available, the mean of intrapersonal distribution $\overline{k_{ij}^L}(\alpha)$ is set to the current value of that image feature $k_{ij}^L(\alpha)$. On the basis of those two measurements, the Gaussian distribution of $P(k_{ij}^L\ (\alpha)|\ c^m)$ is estimated.

After computing $\Delta(c^m|\ k_{ij}^L(\alpha))$ from (7), the discrimination power of node $\boldsymbol{k}_{ij}^L$ for recognition of person c^m is calculated as the geometric average of the discrimination powers of the features $k_{ij}^L(\alpha)$ that comprise node $\boldsymbol{k}_{ij}^L$

$$\Delta\Big(c^m|\boldsymbol{k}_{ij}^L\Big) = \sqrt{\frac{1}{D}\sum_{\alpha=1}^{D}\Delta\Big(c^m|k_{ij}^L(\alpha)\Big)^2} \qquad (8)$$

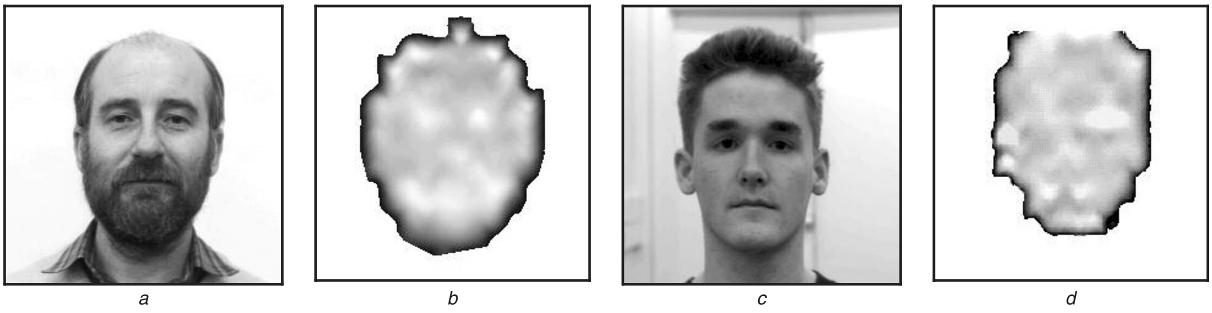

Fig. 4 *Two facial images and corresponding distributions*

a, *c* Facial images
b, *d* Corresponding recognition confidence maps
More salient features are illustrated in light tones

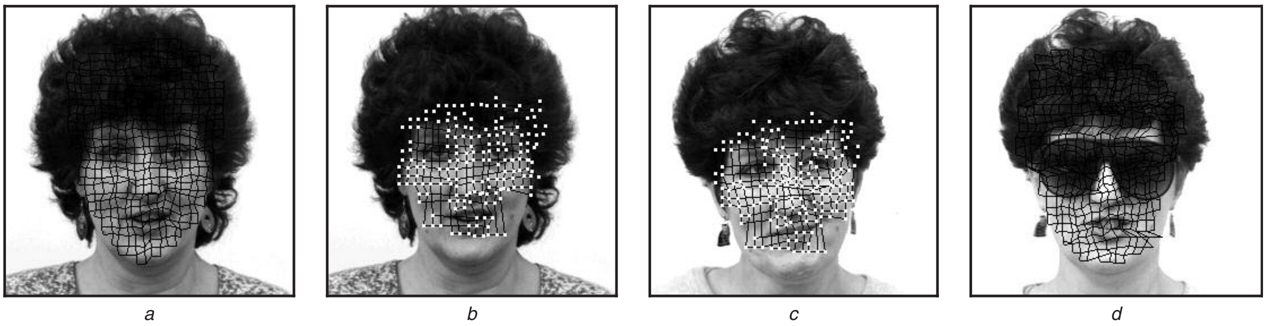

Fig. 5 *Matching of original and sparse model graphs to images*

a Easy test set
b, *c* Test set
d Difficult test set
Images *a* and *d* show matching the high-frequency level of the complete graph; *b* and *c* illustrate matching the high-frequency level of the sparse graph

Fig. 4 illustrates two facial images and corresponding distributions of $\Delta(c^m|\ \boldsymbol{k}^L_{ij})$. The lighter areas in the image distributions Figs. 4*b* and 4*d* correspond to higher discrimination confidence. In general, the areas with higher computed discrimination confidence correspond to the features that the humans would consider as containing characteristic information about the person's identity. Many uncharacteristic features, such as the cheeks, have not been selected as useful for recognition.

The discrimination confidence measure is used as the basis for feature selection (sparsification of the graph) to remove half of the graph nodes with lower discrimination confidence. It is tempting to remove more than half of all the nodes in order to improve the recognition performance while using fewer features. However, as more nodes are removed, the number of loose nodes, which do not have any adjacent neighbours and hence are not attached to the rest of the grid by the spatial links, increases and it has adverse effect on the recognition performance. Future research will be required to preserve spatial correspondence of those loose nodes to the rest of the grid. Meanwhile, sparsification of the graph improves the recognition performance only within a certain range of the sparsification factor.

5 Face recognition performance

The proposed method was tested on the Manchester Face Database [1]. This database contains images of 30 persons and consists of three original sets of images: training set (ten images per person), test set (ten images per person) and a more demanding 'difficult' test set (three images per person). The original intention of the database creators was to use ten images for each person in the database to train the system. As our goal is to train the face recognition system on a training set containing a limited number of

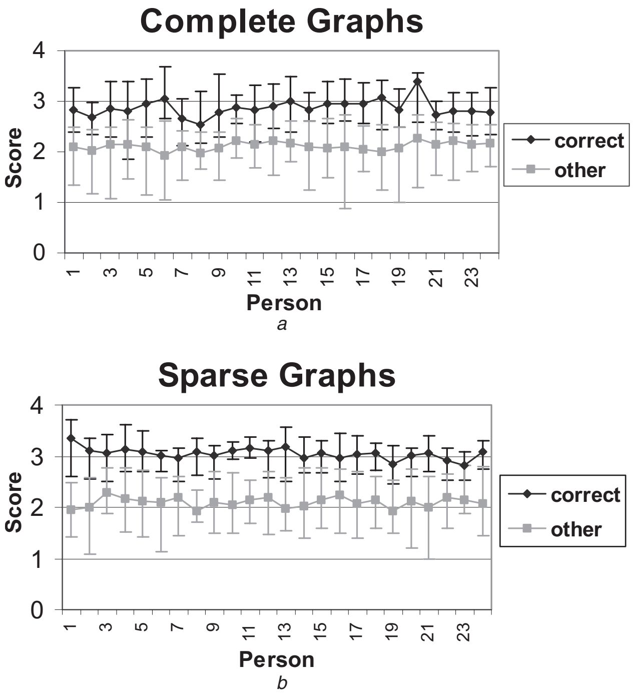

Fig. 6 *Interpersonal and intrapersonal scores*

a Complete graphs
b Sparse graphs

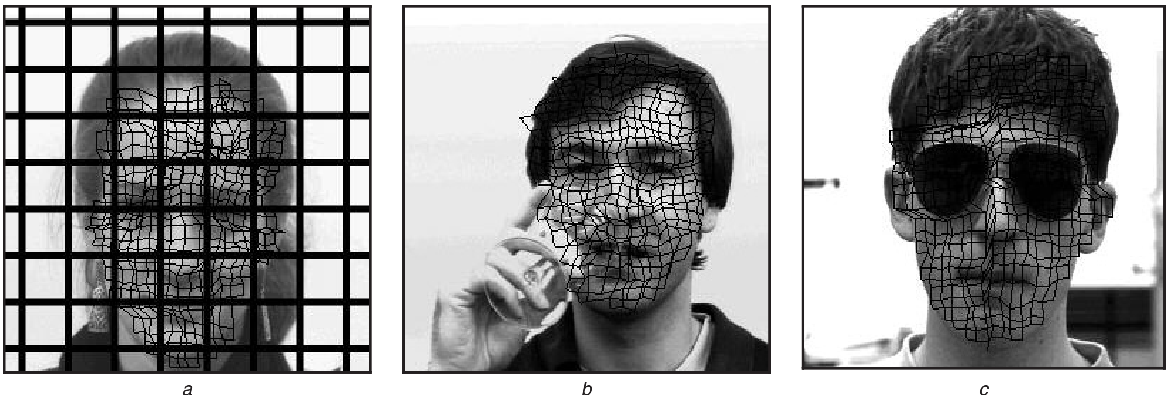

Fig. 7 *Matching complete graphs to images from the difficult test set*

a Image with an overlaid artificial grid
b Image with an occlusion
c Facial image with sunglasses

individuals and use a single image of a person to extract the model graph, the original sets have been rearranged. The new training set consisted of six persons from the training set. The other 24 persons are put in the test set, and a model graph is created from one of the person's image in the original training set. The remaining images from the original training set were considered as easy test images, and the test and difficult test sets of the Manchester Database that were not included in the new training set were used as test and difficult test sets, respectively.

Performance of the system was tested on the newly arranged test, easy test and difficult test sets. Fig. 5 illustrates matching of the high-frequency grid of a person's complete (original) and sparsified model graphs to several other images of the same person. Fig. 5*a* illustrates matching of the original model graph to another image from the easy test set. Fig. 5*b* shows matching of the sparse model graph to the same image. Fig. 5*c* illustrates matching of the sparse graph to an image from the test set, and Fig. 5*d* demonstrates matching of the original graph to an image from the difficult test set. For illustration purposes, the retained nodes in the sparse model graph in Figs. 5*b* and 5*c* are indicated as 3×3 pixel white squares.

Both the original and sparse model graphs achieved 100% performance on the easy test set. Fig. 6 illustrates performance of the original complete (*a*) and sparse (*b*) graphs on the test set. For illustration purposes, the diagrams show matching scores instead of the matching costs. A matching score $\text{Score}_{\text{Total}}$ is computed as

$$\text{Score}_{\text{Total}} = \sum_{L=3}^{5} w_L \times \left(1 - S''_L\left(M^L, I^L(\boldsymbol{x}+\boldsymbol{y})\right)\right) \qquad (9)$$

where each weight w_L indicates a share of each grid in the total matching score. In the figures, the intrapersonal (correct) and interpersonal (other) scores are indicated as the dark and light distributions, respectively. The upper and lower boundaries denote the maximal and minimal scores, and the diamond and square marks illustrate the mean of the intrapersonal and interpersonal scores, respectively. As illustrated in Fig. 6*b*, the gap between the intrapersonal and interpersonal scores increases for sparse graphs. Correspondingly, performance of the sparse graphs on the test set (85%) is better than that of the complete graphs (78%).

The graph-matching procedure was not modified for testing on the difficult test set. The sparse graphs were not tested on the difficult test set, because the proposed feature selection model was not trained to incorporate random occlusions of facial features. Matching the complete graphs on the difficult test set provided 38% recognition rate, defined as the proportion of facial images that are matched to the correct subject. Our graph-matching method performed extremely poorly on images overlaid with a regular grid (Fig. 7*a*). Only one of such images was properly recognised. Certain adaptation of the matching rules might improve the performance on such images. However, such images are artefacts and should not occur in practice very often. On difficult images other than those with overlaid grids, the method performed with a 61% recognition rate. Those images present a more realistic test (see Figs. 7*b* and 7*c* as an example), and the performance on those images is more satisfactory.

The proposed model can be directly compared to an earlier system [1], which utilises three different cues for recognising faces (Table 1). As the earlier system was tested on the original sets of the Manchester Face Database, we can only approximately compare the performance results. Unlike this DLA approach, that system needs to be trained on several images (up to ten) of each person. Moreover, at an 85% successful recognition rate, the proposed model performed better than any single model used by that benchmark system, whose recognition rates were 50.3% using a shape model 78.7% using shape-free grey model and 77.33% using local grey-level models. However, the overall performance of the combined

Table 1: Recognition performance

Set	Our system	Shape model	Shape-free grey-levels	Local grey-levels	Combination
Test	85%[a]	50.3%	78.7%	77.33%	92%
Difficult test	38%[b]	15.6%	31.1%	28.9%	48.9%

[a] When using sparse graphs
[b] When using complete graphs

system was better (92%) [1]. In correctly matching 38% of the images in the difficult test set, the proposed DLA methods also outperformed any single model in the benchmark system for this set, whose face recognition accuracy was 15.6% using a shape model, 31.1% using shape-free grey model and 28.9% using local grey-level models, once again not matching the combination of the three models (48.9%).

6 Discussions and conclusions

The proposed face recognition method was motivated by the observation that perceptual acuity may naturally lead us to select the most discriminative features from a single image of a face, storing these features to identify future variants of the same face among many others that we routinely come into contact with. This premise led to the characterisation of typical intrapersonal and interpersonal facial variations that are later used by a Bayesian decision rule to estimate the discrimination properties of individual facial features from a single image of each subject. The selected localised features are matched to novel facial images using hierarchical elastic graphs.

During the training stage, the distribution of interpersonal variations is computed using only a single image of each person. The intrapersonal variation is obtained from a training set, which contains six people with ten images per person. By relating the intrapersonal distribution to high-level facial features rather than to image coordinates, we store the distribution of intrapersonal variation in the abstract shape-free representation. From this representation, the intrapersonal variations are later generalised for new faces from the test set. In this work, the hair, the nose and the skin region (cheeks, etc.) of the shape-free distribution exhibited larger intrapersonal variations than the eyes and the mouth. We also show the scale dependency of the distribution of intrapersonal variations.

The distribution of the confidence in the identification of facial features, computed using a Bayesian function, illustrates that regions with higher confidence values are consisted with normally accepted salient facial features, surrounding the eyes and the mouth. Nodes with low confidence are removed from the hierarchical graph. It is tempting to leave just a few salient nodes; however, excessive sparseness of the features selected causes the removal of too many of the elastic links, which play an important role in preserving the facial topology in the model graphs.

The localised facial features are matched to new images using hierarchical attributed graphs. A novel feature of the proposed hierarchical graph scheme is the introduction of spatial links between adjacent nodes of each hierarchical level. Another novel idea is a new refinement order, with better matching nodes being allowed larger distortions than other nodes. Those features reduce the computational complexity of matching hierarchical attributed graphs while improving the chances of finding the original correspondence of the nodes.

We show that distortions of higher-resolution grids may be very large compared with the sampling distances of those grids. Further research is needed to establish the relation between distortions of graph links and those of the facial features. In the present implementation, the unclassified distortion information is not included in the matching cost.

The researched method was tested on the Manchester Face Database. The performance of the sparse hierarchical graphs (85%) was better than that of the complete graphs (78%) on the test set. The method outperformed any single face-recognition model used by an earlier system [1]; however, it did not match the combination of all the three models in that system. However, a major advantage of the proposed approach is that once trained it only needs a single image of a new face, whereas the previous system required up to ten training images for each person. A further and potentially significant advantage is the automatic tracking of detailed facial features, which has the potential to be an enabling component for visual tracking in human–computer interaction through affective computation.

7 Acknowledgments

Experiments were conducted using the Manchester Face Database from the University of Manchester, UK. The authors thank the School of Engineering, Liverpool John Moores University for financial support and the General Engineering Research Institute (GERI) for their constant support and encouragement of this work.

8 References

1 Chellappa, R., Wilson, C.L., and Sirohey, S.: 'Human and machine recognition of faces: a survey', *Proc. IEEE*, 1995, **83**, pp. 705–740

2 Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., v.d. Malsburg, C., Würtz, R.P., and Konen, W.: 'Distortion invariant object recognition in the dynamic link architecture', *IEEE Trans. Comput.*, 1993, **42**, pp. 300–311

3 Daugman, J.: 'Face and gesture recognition: overview', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 675–676

4 Lanitis, A., Taylor, C.J., and Cootes, T.F.: 'Automatic interpretation and coding of face images using flexible models', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 743–756

5 Krüger, N.: 'An algorithm for the learning of weights in discrimination functions using a priori constraints', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 764–768

6 Moghaddam, B., and Pentland, A.: 'Probabilistic visual learning for object representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 696–710

7 Picard, R.W., and Klein, J.: 'Computers that recognise and respond to user emotion: theoretical and practical implications', *Interact. Comput.*, 2002, **14**, pp. 141–169

8 Picard, R.W., Vyzas, E., and Healey, J.: 'Toward machine emotional intelligence: analysis of affective physiological state', *IEEE Trans Pattern Anal. Mach. Intell.*, 2001, **23**, pp. 1175–1191

9 Picard, R.W.: 'Toward computers that recognize and respond to user emotion?', *IBM Syst. J.*, 2000, **39**, pp. 705–719

10 Valentine, T.: 'A unified account of the effects of distinctiveness, inversion and race on face recognition', *Q. J. Exper. Psychol.*, 1991, **43A**, pp. 161–204

11 Würtz, R.P.: 'Object recognition robust under translations, deformations and changes in background', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 769–774

12 Daugman, J.: 'Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by 2D visual cortical filters', *J. Opt. Soc. Am. A*, 1985, **2**, pp. 1160–1169

13 Wiskott, L., Fellous, J.M., Krüger, N., and von der Malsburg, C.: 'Face recognition by elastic bunch graph matching', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 775–779

14 Grudin, M.A., Lisboa, P.J.G., and Harvey, D.M.: 'Compact multi-level representation of human faces for recognition'. Proc. 6th Int. IEE Conf. on Image Processing, 1997, Conf. Publ. no. 443, pp. 111–115

15 Field, D.J.: 'Relations between the statistics of natural images and the response properties of cortical cells', *J. Opt. Soc. Am. A*, 1987, **4**, pp. 2379–2394

16 Craw, I., and Cameron, P.: 'Parameterizing images for recognition and reconstruction'. Proc. BMVC 91, Glasgow, 1991, pp. 367–370