Face Recognition by the Construction of Matching Cliques of Points

Fred Stentiford ; University College London, Electronic & Electrical Engineering Dept, Gower St, London, UK

Abstract

This paper addresses the problem of face recognition using a graphical representation to identify structure that is common to pairs of images. Matching graphs are constructed where nodes correspond to local brightness gradient directions and edges are dependent on the relative orientation of the nodes. Similarity is determined from the size of maximal matching cliques in pattern pairs. The method uses a single reference face image to obtain recognition without a training stage. Results on samples from MegaFace obtain a 100% correct recognition result.

Introduction

The use of intuitively plausible features to recognise faces is a powerful approach that yields good results on certain datasets. Where it is possible to obtain a truly representative set of data for training and adjusting recognition parameters, optimal performance can be attained. However, when facial images are distorted by illumination, pose, occlusion, expression and other factors, some features become inappropriate and contribute noise to the discrimination on unseen data. Indeed it can never be known in advance what distortions will be present in unseen and unrestricted data and so features that are applied universally are likely to reduce performance at some point.

Many approaches to face recognition are reported in the literature [1,2]. Graph matching approaches provide attractive alternatives to the feature space solutions in computer vision. Identifying correspondences between patterns can potentially cope with non-rigid distortions such as expression changes, pose angle and occlusions. However, graph matching is an NP-complete problem and much of current research is aimed at solving the associated computational difficulties.

SIFT feature descriptors are used by Leordeanu et al [3] to construct spectral representations of adjacency matrices whose nodes are feature pair correspondences and entries are dependent on feature separations. Objects in low resolution images are recognised by matching correspondences against a set of pre-trained models. Felzenszwalb et al [4] also match a graphical model of specific objects to images in which parts are matched according to an energy function dependent on colour difference and relative orientation, size and separation. Fergus et al [5] avoid the computational complexity of a fully connected shape model by adopting a "star" model that uses "landmark" parts. The model is trained using specific feature types and recognition is obtained by matching appearance densities of model parts. Kim et al [6] reduces the computational demands by first segmenting one of the images. Each region is mapped using SIFT descriptors and a function dependent on distortion, ordering, appearance and displacement is minimised to obtain appropriate candidate points and region correspondence.

A more general approach by Duchenne et al [7] uses graph matching to encode the spatial information of sparse codes for pairs of images. An energy function is maximised using a graph cuts strategy that is dependent on node feature correlation, reduced node displacement and discouraging node crossing. Duchenne et al [8] also uses a tensor based algorithm to match hypergraphs in which correspondences are identified between groups of nodes and hyperedges linking them. The method is illustrated by matching two similar faces using triples of SIFT descriptors. Celiktutan et al [9] also match hypergraphs connecting node triples in the spatial-temporal domain by minimizing an energy function. Computation is reduced by considering a single salient point in each video frame and limiting connections along the time dimension.

Kolmogorov et al [10] present a graph-cut algorithm for determining disparities that ensures that single pixels in one image are assigned single pixels in the second image and occlusions are handled correctly. An energy function is employed that is minimised by reducing the intensity difference between pixels, by penalizing pixel occlusions, and requiring neighbouring pixels to have similar disparities.

Berg et al [11] sets up correspondences by identifying edge feature locations and measuring their similarity by using the correlation between feature descriptions and the distortion arising from local changes in length and relative orientation. An approximate version of Integer Quadratic Programming is used to detect faces. Cho et al [12] proposes a method for growing matching graphs where nodes represent features and edges the geometric relationships. The Symmetric Transfer Error is used to measure the similarity of node pairs and the reweighted random walk algorithm to match nodes.

Shape driven graphical approaches [13-15] including active appearance models assign fiducial points to nodes and maximise a similarity function to obtain recognition of candidate images. Deep neural networks [16] obtained a 97% result on 17 facial images from the Yale Database A using facial features and the remaining 148 images for training the network.

This paper makes use of a fully connected graph matching representation in order to measure the similarity of pairs of patterns. Nodes are pixels and edges take the value of the relative orientation of the two pixels. Node values are assigned the value of brightness gradient direction at that locality. Patterns represented by graphs being compared match if the pair obtains the highest number of matching node values and edge values. This simple framework has the advantage that the graph matching process is much faster because it can be reasonably assumed that if local node edges match, the relative orientation of many distant nodes will not vary significantly within that locality and will therefore also match. No training data is required and no other features are employed. The method requires single reference face images from each of the individuals to be recognised. Ideally the reference faces images should be of high quality with the background removed to improve performance. There are no such constraints on the candidate face images. Results using an earlier related approach [17] obtained a 100% correct result on the Yale Face Database A [18].

Figure 1. Ten of the 100 MegaFace candidates
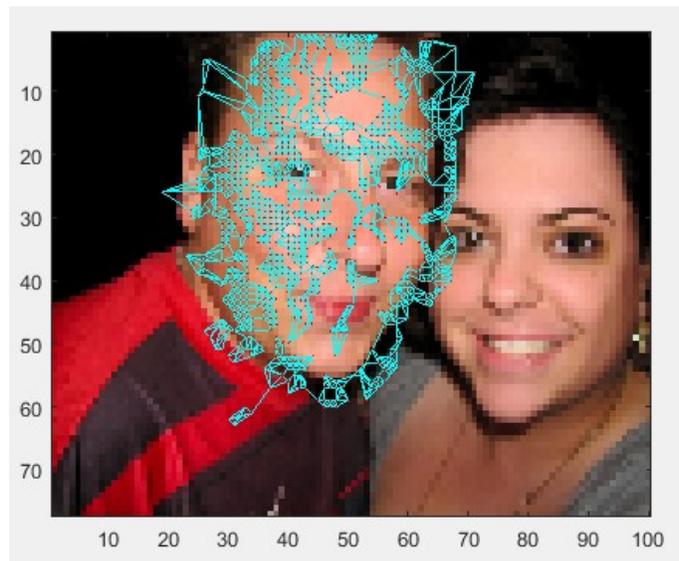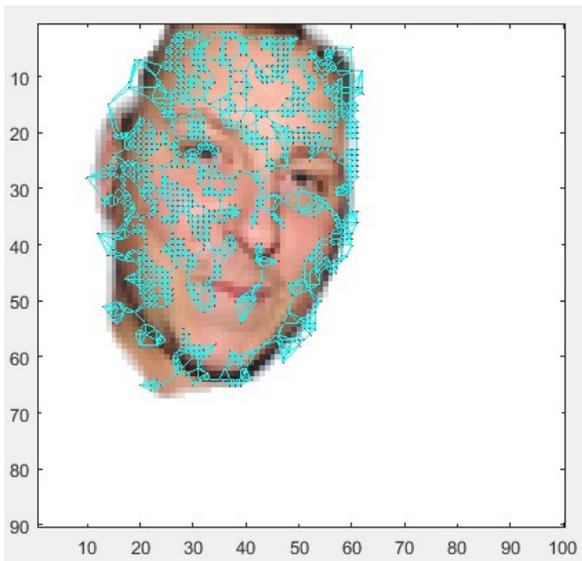

Figure 2. Corresponding reference faces

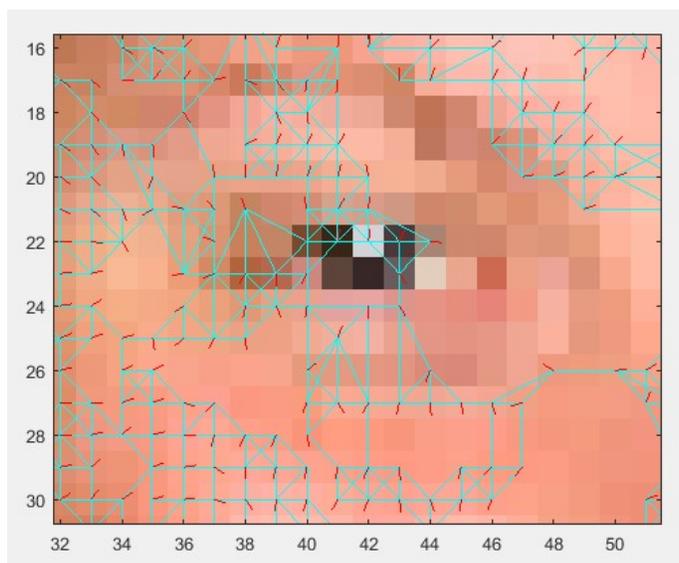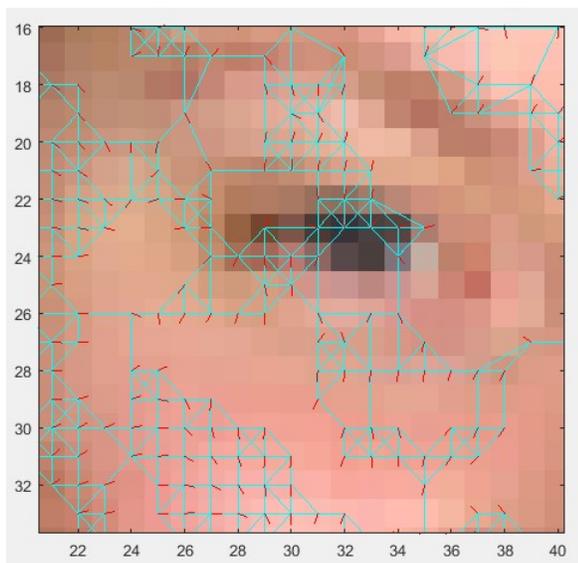Figure 3. Reference face matched against candidate

Figure 4. Reference and candidate clique magnified

Proposed Approach

The approach taken in this paper detects structure that is common between pairs of images and uses the extent of such structure to measure similarity. In this case the size of the largest structure found to match both patterns is the number of nodes in the corresponding fully connected maximal graph or clique.

A pictorial structure is represented as a collection of parts and by a graph $G = (V, E)$ where the vertices $V = \{v_1, \dots, v_n\}$ correspond to the parts and there is an edge $(v_i, v_j) \in E$ for each pair of connected parts v_i and v_j. An image part v_i is specified by a location x_i. In this paper parts v_i correspond to individual pixels. Given a set of vertices $V^1 = \{v_1^1, \dots, v_n^1\}$ in image 1 that correspond to a set of vertices $V^2 = \{v_1^2, \dots, v_n^2\}$ in image 2 the following conditions are met by all parts to form a clique

$$\left| d_g(x_i^1) - d_g(x_i^2) \right| \le \varepsilon_1 \tag{1}$$

$$\left| d_a(x_i^1, x_j^1) - d_a(x_i^2, x_j^2) \right| \le \varepsilon_2 \quad \forall i, j \quad i \ne j \tag{2}$$

where $d_g(x_i)$ is the grey level gradient direction at x_i and $d_a(x_i, x_j)$ is the angle subtended by the point pair (x_i, x_j). Clique generation begins with the selection of a random pair of pixels (x_i^1, x_j^1) from reference image 1 and a pair (x_i^2, x_j^2) from candidate image 2 that satisfy (1,2). A new pair of points (x_k^1, x_k^2) is added where

$$\left| d_g(x_k^1) - d_g(x_k^2) \right| \le \varepsilon_1 \tag{3}$$

$$\left| d_a(x_k^1, x_m^1) - d_a(x_k^2, x_m^2) \right| \le \varepsilon_2 \tag{4}$$

where x_k^1 has not already been selected and x_m^1 is the closest point to x_k^1 from those already selected from reference image 1

$$m = \arg\min_p \left| x_p^1 - x_k^1 \right|$$

New candidate points (x_k^1, x_k^2) are selected randomly and added to the clique if conditions (3,4) are satisfied. Up to N attempts are made to find a new point after which the current clique is completed and the construction of a new clique started. The search proceeds on a trial and error basis and the selection is not guided by additional heuristics as these have always been found to damage performance. After the generation of P cliques the largest is retained. Let the number of nodes in the maximal clique extracted between the reference image for class c and candidate image i be n_c^i. The classification of image i is given by C_i where

$$C_i = \arg\max_c n_c^i$$

The relationship between points is not dependent upon their separation or absolute position and therefore the similarity measure is translation and scale invariant. It also means that there is no special constraint placed on the disparity of points that is dependent on their separation. The measure is partially invariant to the rotation of the images to within the angle ε_2. It should also be noted that although the cliques are maximal in terms of the algorithm, there is no guarantee that the cliques extracted are the largest theoretically possible; the solution of an NP-complete problem would be necessary to confirm this.

MegaFace Database

In order to make a further assessment of the performance of the latest algorithm, 100 reference faces and 100 candidate faces were taken from the MegaFace MF2 Training Dataset [19]. 10 of the 100 candidate faces are shown in Fig. 1 and the 10 corresponding reference faces are shown in Fig. 2. The reference faces have had the background set to white because commonality with the background adds noise to the result. In addition all the images have been linearly scaled down to be 100 pixels wide before analysis.

The threshold on the brightness gradient direction is $\varepsilon_1 = 55°$. The threshold on the angular difference between matching pairs of points in each image is $\varepsilon_2 = 20°$. Up to N=100 attempts are made to add new points to a clique and P=100 cliques are generated for each image i, the maximal clique identified, and the classification C_i determined. This defined a fixed framework for clique extraction but with two very broad thresholds thereby enabling more points to become candidates for inclusion in a clique. There is therefore less emphasis placed on the information possessed by individual pixel properties than that contained in the structural relationships between the points forming the clique.

A 100% result was obtained in which all candidate faces obtained the highest matching score or the largest clique of nodes with the correct reference image. Figure 3 shows the maximal matching clique on the reference and candidate images. There are 1053 nodes. Figure 4 shows magnified matching sections from both images which illustrates the directions of the brightness gradients of each node with red lines. The shape of the clique is similar but differences are allowed within the specified thresholds on angular differences and brightness gradient directions.

Discussion

The method obtained a 100% correct result largely because the face references were visually similar to the correct candidate and visually different to other candidates. Errors will arise with lower quality images and images from individuals that are visually similar.

The similarity measure used in this paper identifies structure that is common to pairs of patterns. This means that pattern classes can be accommodated that contain patterns that only possess features in common with just some of the other class members. Some patterns can therefore be members of the same class but have no features in common, a situation that is not permitted by several feature based approaches that are dependent on a metric.

The two parameters, brightness gradient direction and relative angular difference ($\varepsilon_1, \varepsilon_2$) that define the operation of the clique matching process are independent of the pattern content and were unchanged when processing faces. A possible link with natural vision is illustrated in a paper by Potvigin [20] who identified neurons in the in the visual system of the cat that were sensitive to the orientation of brightness gradients. Increasing the number of cliques generated (P) and the number of attempts to add new nodes

to a clique (N) increases the likelihood of discovering larger cliques. However, currently the search for maximal cliques automatically obtains registration and P and N therefore could be reduced (with the associated reduction in processing) given prior information on the registration of the pattern pairs.

The clique extraction process is scale independent as is illustrated by the flexible matching in Fig. 4. The recognition is also independent of brightness. Illumination generates shadows but within the shadow the grey level gradient direction is not affected. The edges of shadows do prevent matching but only in small regions where the shadow edge is positioned.

Earlier research on modelling visual attention in the human visual system has employed maximal cliques to determine the similarity of regions within the *same* pattern [21]. This has enabled the background in an image to be recognised thereby isolating and identifying salient objects.

The computation takes approximately 25 seconds in Matlab to compare two grey level images. This time is much less for face pairs that differ significantly because searches are abandoned while cliques are still small, but is larger if the images are very similar. This may be reduced as discussed above with prior registration. However, the extraction of each clique is an independent operation and may therefore be conducted in parallel both for each reference c and the P attempts at maximal clique construction. Furthermore parallel operations could also be introduced during graph matching itself by allowing additional nodes to be added simultaneously at different locations. This is possible because the conditions for addition are dependent only on local node properties. This means that the overall potential for a speedup of many orders of magnitude is possible in an appropriate implementation.

Conclusions

This paper has demonstrated the existence of matching clique-like structures that obtain no errors when applied to the classification of a selection of images from MegaFace. The approach requires no training stage or pre-selected features. Although the serial implementation is slow there is potential for very fast parallel operation.

Support for the approach is also obtained from its use in modelling aspects of human vision. Further work is necessary on larger and more challenging datasets. There is scope for including colour in the clique node properties as well as increasing the image resolution to improve discrimination on more detailed images.

References

[1] Patel, R., Rathod, N. and Shah, A., "Comparative analysis of face recognition approaches a survey," Int. J. of Computer Applications, 57(17) 50-61 (2012).

[2] Naruniec, J., "A survey on facial features detection," Int. J. of Electronics and Telecommunications, 56(3) 267-272 (2010).

[3] Leordeanu, M. and Hebert, M., "A spectral technique for correspondence problems using pairwise constraints," ICCV, (2005)

[4] Felzenszwalb, P. F. and Huttenlocher, D. P., "Efficient matching of pictorial structures," CVPR, (2000).

[5] Fergus, R., Perona, P. and Zisserman, A., "A sparse object category model for efficient learning and exhaustive recognition," CVPR, (2005).

[6] Kim, J. and Grauman, K., "Asymmetric region-to-image matching for comparing images with generic object categories," CVPR, (2010).

[7] Duchenne, O., Joulin, A. and Ponce, J., "A graph-matching kernel for object categorization," ICCV, (2011).

[8] Duchenne, O., Bach, F., Kweon, I. and Ponce, J., "A tensor-based algorithm for high-order graph matching," IEEE Trans PAMI, 33(12) 2383-2395 (2011).

[9] Celiktutan, O., Wolf, C. and Sankur, B., "Fast exact matching and correspondence with hyper-graphs on spatio-temporal data," Technical Report LIRIS-RR-2012-002, INSA-Lyon, (2012).

[10] Kolmogorov, V. and Zabih, R., "Computing visual correspondence with occlusions using graph cuts," ICCV, (2001).

[11] Berg, A. C., Berg, T. L. and Malik, J., "Shape matching and object recognition using low distortion correspondences," CVPR, (2005).

[12] Cho, M. and Lee, K. M., "Progressive graph matching: making a move of graphs via probabilistic voting," CVPR, (2012).

[13] Wiskott, L., Fellous, J-M., Kruger, N. and von der Malsburg, C., "Face recognition by elastic bunch graph matching," IEEE Trans. Pattern Anal. Machine Intell., 16, 775-779, (1997).

[14] Cootes, T. F., Edwards, G. J. and Taylor, C. J., "Active Appearance Models," IEEE Trans. Pattern Anal. Machine Intell., 23, 681-685 (2001).

[15] Matthews, I. and Baker, S., "Active appearance models revisited," Int. J. Computer Vision, 60(2) 135-164 (2004).

[16] Gupta, P., Saxena, N., Sharma, M. and Tripathi, J., "Deep Neural Network for Human Face Recognition," Int. J. of Engineering & Manufacturing, 8(1), 63-71 (2018).

[17] Stentiford, F. W. M., "Face Recognition by Detection of Matching Cliques of Points," Image Processing: Machine Vision Applications VII, Electronic Imaging Conf., San Francisco, 2 - 6 Feb. (2014).

[18] Yale Face Database Database, http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[19] MegaFace Database, http://megaface.cs.washington.edu/.

[20] Potvigin, N. F., Poeppel, E., Kiseleva, N. B., Kozlov, I. V., Vershinina, E. A., and Granstrem, M. P., "The sensitivity of neurons in the lateral geniculate body of the cat to the orientation vectors of brightness gradients," Neuroscience & Behavioural Physiology, 31(6) 657-668 (2001).

[21] Stentiford, F. W. M., "Saliency identified by absence of background structure," Human Vision and Electronic Imaging XVIII Conf., San Francisco, 3 - 7 Feb. (2013).

Author Biography

I left school with an open scholarship in mathematics to St Catharine's College, Cambridge obtaining an honours degree in mathematics. I was awarded a PhD from the Electronics Department of the faculty of Engineering and Applied Science at Southampton University. I joined British Telecom and led a group researching new ways of analysing, synthesising, and delivering content. In April 2002 I accepted a chair in the Electrical & Electronic Engineering department of University College London. Since 2009 a new research direction has widened applications to visual attention,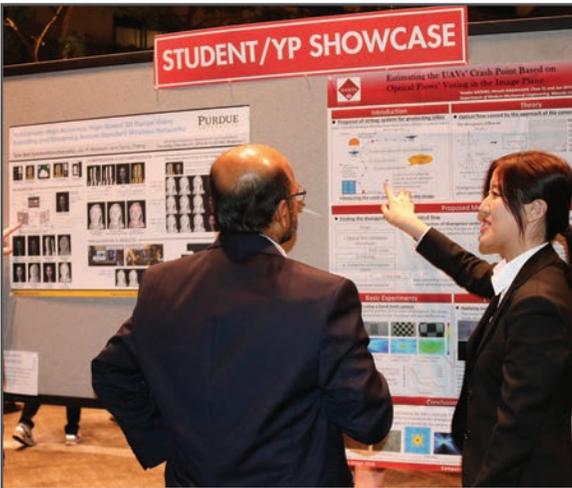 face recognition and text similarity.