Finding Face Features *

Ian Craw, David Tock, and Alan Bennett

Department of Mathematical Sciences, University of Aberdeen, Aberdeen AB9 2UB, UK.

Abstract. We describe a computer program which understands a greyscale image of a face well enough to locate individual face features such as eyes and mouth. The program has two distinct components: modules designed to locate particular face features, usually in a restricted area; and the overall control strategy which activates modules on the basis of the current solution state, and assesses and integrates the results of each module.

Our main tool is statististical knowledge obtained by detailed measurements of many example faces. We describe results when working to high accuracy, in which the aim is to locate 40 pre-specified feature points chosen for their use in indexing a mugshot database. A variant is presented designed simply to find eye locations, working at close to video rates.

1 Introduction

We describe a program to recognise and measure human facial features. Our original motivation was to provide a way of indexing police mugshots for retrieval purposes. Another use comes when identifying points on a face with a view to cartooning the face [2] or to obtain a more useful Principal Component Analysis of face images [3], leading to a compact representation suitable for matching. The work on blink rate is part of a PROMETHEUS (CED2 - Proper Vehicle Operation) study by the Ford Motor Company on driver awareness, for which blink rate is a useful indicator. The aim is to correlate the blink rate, obtained during normal driving conditions, with the output from other sensors. Ideally this would enable blink rate to be predicted from measurements which can be made more cheaply as part of the normal sensor input available from a modern car.

The system aims to locate a total of 40 feature points within a grey-scale digitized full face image of an adult male. We initially choose to ignore glasses and facial hair, although such images are occasionally used to test the robustness of the system. The points chosen are those described in Shepherd [8]; thus allowing us to utilise data originally recorded by hand. A total of 1000 faces were measured, and the locations of the 40 points on each face were recorded. This data has been normalized and forms the basis of the model described below. Identification is confirmed by overlaying the points on the image; the points are usually linked with straight lines to form a *wire frame* face as shown in figure 1.

Our system is in two distinct parts:

- a number of independent recognition modules specialised to respond to individual parts of the face; and

* This research was supported by a United Kingdom Science and Engineering Research Council project grant number GR/E 84617 to Ian Craw and J Rowland Lishman (Computing Science, Aberdeen). DT was funded by this project; AB was supported by an SERC studentship. We thank Roly Lishman for many valuable discussions on faces.

– a control structure, driven by a high level hypothesis about the location of the face, which invokes the feature finding modules in order to support or refute its current hypothesis.

The program endeavours to confirm that a face is located within the image in a two part process. Possible coarse locations for the face are sought *ab initio* providing *contexts* for further search, using modules capable of providing reliable, although not necessarily accurate locations even when given a wide search area. Contexts are then refined and assessed. In this phase, feature finding modules are usually called with very restricted search areas, determined by statistical knowledge of the relative positions of features within the face. When all the required feature points are located in a single context, this identifies the face itself; and the features' locations provide mutual support.

2 Feature Experts

Feature experts obtain information directly from the image. Our current set range from simple template matchers to much more elaborate deformable template models. Implementing a template matcher is trivial, and execution is rapid, but problems arise from their heavy dependence on scale and orientation, and multiple responses from many parts of the image: however FindFace is normally confirming a good working hypothesis; our templates are generated dynamically, tuned to the expected size of the feature, and applied to a small search area.

We describe methods designed primarily for initial location as "global methods" as opposed to the "local methods" used to assess or verify a location proposed by the existing context. So, for example the global methods for locating a single eye look either for dark, compact blobs completely surrounded by lighter regions or for areas of the image with a substantial high frequency component. Locally the eyes use a probabilistic eye locater rather like the outline finder we describe below, or even the blob detector confined to a small area, if the uncertainty of the location is still high. A full list of feature experts can be found in [4]

A number of algorithms have been proposed for finding the outline of a head, dating at least from Kelly [6]. Our approach is inspired by the work of Grenander et al.[5] and [7], in which a polygonal template outline is transformed at random to fit the data, governed by our detailed statistics [1]. The advantage of this approach is that the background can be cluttered (cf figure 3), and the initial placing of the outline is not required to be *outside* the head.

The approximate location, scale and orientation of the head is found by repeatedly deforming the whole template at random by scaling, rotation and translation, until it matches best with the image whilst remaining a feasible head shape. This feasibility is determined by imposing statistical constraints on the range of allowable transformations. The optimisation, in both stages, is achieved using simulated annealing; although this means the method is not rapid, it appears to be particularly reliable.

A further refinement is then achieved by transforming the individual vectors within the polygon under certain statistical constraints. Consider the outline as a list of vectors $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ with $\mathbf{v}_i \in \mathbb{R}^2$ for each i, where the representation is obtained by regarding each vector as based at the head of the previous one in the list. Since the outline is closed. $\mathbf{v}_1 + \ldots + \mathbf{v}_n = 0$; a new outline may be generated by applying $(n-1)$ transformations from the group $O(2) \times US(2)$, where O is the orthogonal group and US is the group of uniform scale change, to the first $n-1$ vectors in the list. We represent elements of this

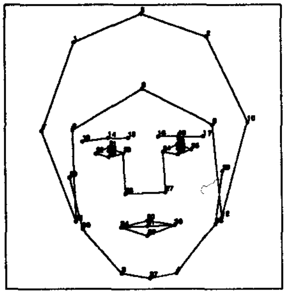

Fig. 1. *Wire frame model.*

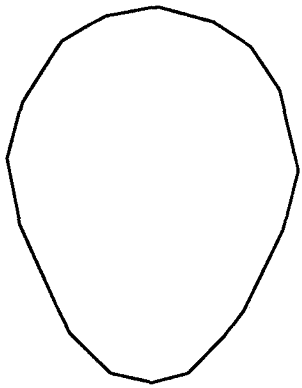

Fig. 2. *The template for a head outline.*

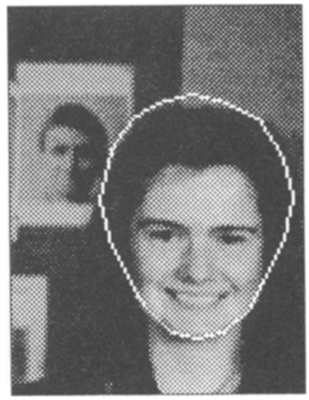

Fig. 3. *Identification in a cluttered image.*

group as matrices of the form

$$\begin{pmatrix} u & v \\ -v & u \end{pmatrix}.$$

When $u = 1$ and $v = 0$, each transformation is the identity map, and the resulting outline is the (initial) average one. Our first approximation to generating a variable outline is to chose $u_i \in \mathcal{N}(1, \sigma_u)$ and $v_i \in \mathcal{N}(0, \sigma_v)$ for $1 \leq i \leq n-1$, where the corresponding variances are calculated from our detailed measurements. This then gives a shape (behaviour) score of the form

$$B = \exp\left(-\frac{1}{2}\left(\frac{1}{b_u}\sum_{i=0}^{n-1}\frac{(u_i-1)^2}{\sigma_{u_i}^2} + \frac{1}{b_v}\sum_{i=0}^{n-1}\frac{v_i^2}{\sigma_{v_i}^2}\right)\right).$$

where the constants b_u and b_v are associated with *independent* behaviour at each site.

This alone allows too much uncoordinated variation between neighbouring vectors. To ensure more coordinated, head-like, polygons, we also place a measure on the change in transformations generating adjacent vectors, making the assumption that the u's and v's are realisations of Markov random chains of order 1. This gives a component of the shape (acceptance) score in the form

$$A = \exp\left(-\frac{1}{2}\left(a_u\sum_{i=0}^{n-1}\frac{(u_i-u_{i-1})^2}{\sigma_{u_i}^2} + a_v\sum_{i=0}^{n-1}\frac{(v_i-v_{i-1})^2}{\sigma_{v_i}^2}\right)\right)$$

where the constants a_u and a_v scale the bonding relations.

By varying the parameters a_u, a_v, b_u and b_v, we control the variability and coordination of the transformations. High values of a_u and a_v force coordination, whilst high values of b_u and b_v allow greater variability. The σ's represent the variability of particular vector transformations and allow variability in the distributions from vector to vector. Our description means that the first point of the outline will remain fixed; in fact we apply the above procedure only to a sublist of the original outline list (a sweep site in Grenander's terminology). Since we regard the list as cyclic, the base point can move, and the outline slowly change location.

3 Control

A *feature* is either a feature point, or recursively, a suitable grouping of features or feature points; natural groupings include the mouth, consisting of five feature points, the eyes and of course the whole face itself. As a putative location for each new feature is returned by the model feature expert, the best affine transformation between face and image co-ordinates is calculated using a least squares fit of the located features and the mean model feature locations. The new location is only accepted if the residual in this matching is reduced with the additional point.

Fig. 4. *Initial context destined for rejection in favour of the one in figure 5.*

Fig. 5. *Initial context for which refinement succeeds.*

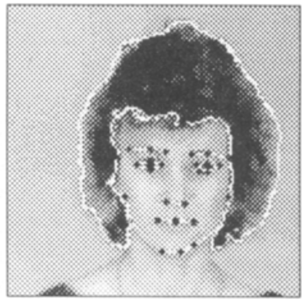

Fig. 6. *Final result obtained for the context shown in figure 5.*

The model expert is responsible for creating an initial set of contexts Two such contexts are shown diagramatically in figures 4 and 5. We deliberately generate a number of contexts at this stage; ultimate success requires a context to be relatively close to the correct position; in practice we rarely need refine more than three contexts.

The location of the remaining features and subsequent refinement of all features forms the second phase of the operation of FindFace. The feature's location in the model is transformed into image coordinates; a search area is also defined based on the model feature's variance and the current context residual. Each feature expert in the list is consulted in turn until one returns a positive result. Since the residual decreases monotonically apart from when each of a finite number of feature points is added to the context, convergence necessarily occurs; in fact it happens quite quickly. A completed context is shown in figure 6.

4 Results

The FindFace system has successfully been demonstrated on many interested visitors to the department, and their faces often include attributes FindFace is not designed to work with — glasses, beards or females. More rigorous testing has been performed on random batches of 50 images from our library of faces, including subjects with beards and glasses. Another test used a sequence of 64 images of a moving subject, at about 8 frames per second. In a typical batch,

- the head position is correctly located in all images, with the outline completely detected in 43 cases — the region normally missing from the remainder is the chin;

- the absence of feature experts for the eyebrows reduces the number of possible feature point locations to 1462, of which the system claims to identify 1292;
- of the located points, 6% were inaccurately or incorrectly identified — again the mouth and chin region were usually in error.

The problem with the mouth and chin region is partially attributable to the inclusion of subjects with beards and moustaches; somewhat surprisingly, glasses do not interfere as much as originally anticipated. On the sequence of 64 images, the overall success rate increased to greater than 95%. This result was achieved by processing each image *ab initio*; better results would have been obtained by using a priori knowledge obtained from the previous image(s).

5 Working Faster

A second implementation based on the same design philosophy aims to make detailed measurements of the eye region from a real time video sequence, and hence detect eyelid separation and so blink rate. The other points in the model need only be located for corroboration purposes; valuable since the image sequence originates from a camera mounted on the dashboard of a car, pointing towards the driver. The resulting images suffer from poor contrast, and a variety of different noise elements caused by vehicle motion, changing lighting conditions etc. We now incorporate facilities for initialising the system with a new subject, and tracking movement between frames. A system has been successfully demonstrated that tracks the eye movement at approximately 5 frames per second, although the detailed eye measurements were not being produced.

References

1. A. D. Bennett and I. Craw. Finding image features using deformable templates and detailed prior statistical knowledge. In P. Mowforth, editor, *British Machine Vision Conference 1991*, pages 233–239, London, 1991. Springer Verlag.
2. P. J. Benson and D. I. Perrett. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, 3(1):105–135, 1991.
3. I. Craw and P. Cameron. Parameterising images for recognition and reconstruction. In P. Mowforth, editor, *British Machine Vision Conference 1991*, pages 367–370, London and Berlin, 1991. British Machine Vision Association, Springer Verlag.
4. I. Craw, D. Tock, and A. Bennett. Finding face features. Technical Report 92-15, Departments of Mathematical Sciences, University of Aberdeen, Scotland, 1991.
5. U. Grenander, Y. Chow, and D. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes*. Research Notes in Neural Computing. Springer-Verlag, New York, 1991.
6. M. Kelly. Edge detection in pictures by computer using planning. In B. Meltzer and D. Michie, editors, *Handbook of research on face processing*, pages 397–409. Edinburgh University Press, Edinburgh, 1971.
7. A. Knoerr. Global models of natural boundaries: Theory and applications. Pattern Analysis Technical Report 148, Brown University, Providence, RI, 1988.
8. J. W. Shepherd. An interactive computer system for retrieving faces. In H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. Young, editors, *Aspects of Face Processing*, chapter 10, pages 398–409. Martinus Nijhoff, Dordrecht, 1986. NATO ASI Series D: Behavioural and Social Sciences - No. 28.

This article was processed using the LaTeX macro package with ECCV92 style