A Knowledge Based System for Measuring Faces — **Source link** ⬈

David Tock, Ian Craw, Roly Lishman

Related papers:

- Eigenfaces for recognition

- A Knowledge Based Approach for Image Understanding

- Knowledge-based machine vision

- A knowledge-based videotheodolite measurement system for object representation/monitoring

- Some Aspects Concerning Geometric Forms Automatically Find Images and Ordering them Using RobotStudio Simulation

A Knowledge Based System for Measuring Faces

David Tock, Ian Craw and Roly Lishman

Departments of Mathematical Sciences and Computing Science *
University of Aberdeen
Aberdeen AB9 2TY, Scotland

Recognising faces is now an attractive computer vision task. Methods used for machine parts; recognizing straight lines or circular arcs, and the use of detailed geometric models, are not directly relevant, yet with many human faces to recognize, it is worthwhile building a system solely for this task. The system described here uses a blackboard architecture and aims to obtain a set of measurements from an image of a face. A key feature is the use of knowledge about faces, including a statistical knowledge base derived from detailed measurements of a large number of sample images.

Our original motivation for considering faces was to provide a way of indexing police mugshots for retrieval purposes. Our images are thus in a fairly standard form, and are of good quality. However, the ease with which we recognise a face belies the complexity of the task. For example figure 1(a-c) shows three images immediately recognizable as faces, but which have little in common. Furthermore, even at very low resolutions, faces are still recognizable (figure 1d).

In fact, the system we describe does not attempt to recognize individual faces explicitly, rather it aims to locate a number of feature points within the face. These feature points are then used to perform the recognition stage, or to help retrieve a desired image.

Our starting point is that individual features can frequently be identified using low level vision techniques. However a serial approach, finding features in some predetermined order, gives rise to a combinatorial explosion. This can be avoided by pruning possible identifications using inter-relationships between the features, but catastrophic failure results when a successful identification is rejected at too early a stage in processing. An early approach along these lines was by Sakai et al [12], [13] and Kanade [9]. A later study was by Craw et al [6].

Reasoning with a high level representation of the problem can also be necessary. Consider the difficulty of accurately locating the significant feature point shown in figure 2, an enlarged portion of figure 1a. Although the faces in figure 1 were easy to understand, a small area
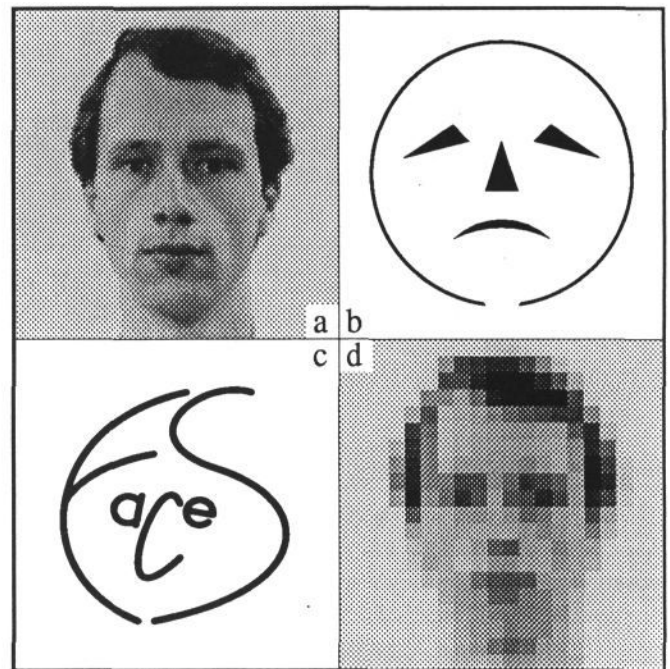
Figure 1: Examples of very different images, all recognizable as faces.

of a digitized image taken out of context may be much more difficult. Obviously, the semantics of a larger proportion of the image are required. Yet we need to work to a high degree of accuracy; in our images the difference in feature locations between people may well be only a few pixels.

In this paper we describe a blackboard based system, GetFace, currently being implemented which combines a number of different types of knowledge in order to locate the feature points. We focus on two of the modules or knowledge sources, and presents some initial results.

PROBLEM SPECIFICATION

Our system aims to locate accurately a total of 37 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 images can be at an arbitrary resolution; the system starts with a low resolution image, and progressively tracks features through
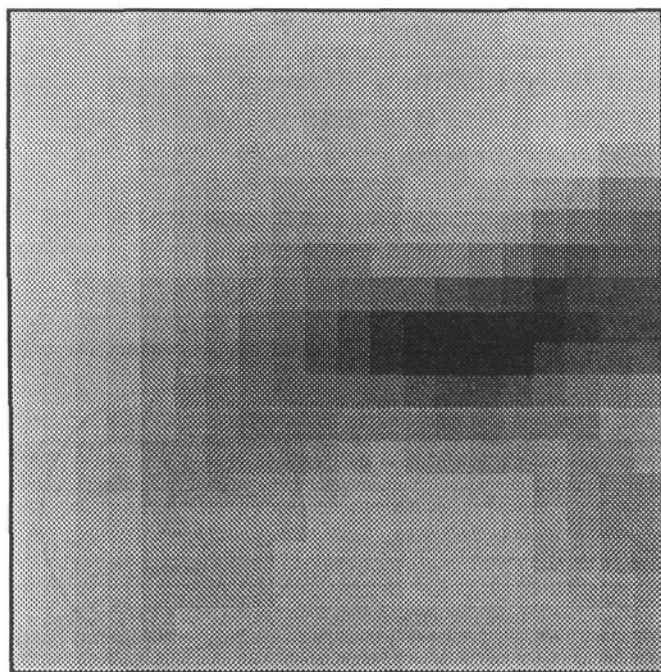
Figure 2: **Magnified view of part of figure 1a.**

increasing resolutions until the original is reached.

The points chosen are those described in Shepherd [15]. A number of different sets of such points have been proposed (for example Goldstein et al [1], Kanade [9]), and our system could easily be adapted to work with these; however our choice allows us to utilise data originally recorded by hand. A total of 1000 faces were measured, and the locations of the 37 points on each face were recorded. This data has been normalized and forms the basis of the model expert described below. For display purposes, the points are usually linked with straight lines to form a *wire frame* face as shown in figures 4 and 5.

THE BLACKBOARD ARCHITECTURE

GetFace uses a blackboard architecture; a method of organizing independent knowledge based methods (dubbed experts) with different competencies, and coordinating their execution. Experts communicate via a common data area, the blackboard, contributing to a solution that develops incrementally and opportunistically. For a full description see Englemore and Morgan [7].

Our system, based on SBS [2], and implemented primarily in POP11 utilises a frame-like structure [17] for representing data on the blackboard. Both declarative and procedural knowledge may be incorporated in experts, along with calls to functions written in other languages, particularly useful for computationally intensive operations.

No predetermined order of processing is imposed by the system. Experts monitor the state of data on the black-board; when particular conditions are met they perform some action, usually modifying the blackboard in the process. An important feature of the experts in our system is that they all provide a measure of their confidence in their results.

A control expert monitors the system during execution; it is able to resolve conflicts that develop between experts, and provides an interface to the user for interactively testing individual subsystems. Its main function however, is to coordinate the execution of experts and hence guide the conversion of data on the blackboard between different levels of representation.

Although currently implemented for only a subset of the features required, early results indicate that the system performs as expected. The addition of further feature experts is facilitated by the existence of a functioning subsystem.

KNOWLEDGE SOURCES

The overall arrangement of the system is shown in (figure 3), which illustrates the division of knowledge into categories:

- general properties of faces,
- individual features within faces,
- image processing programs,
- control knowledge.

A consequence of having such well defined input images is that an unusually high level of top down reasoning can take place. During initialisation a single expert can usually perform the now trivial task of locating the approximate position of a face within an image. This provides a context for the other experts to perform the main task, that of refining the location of the feature points.

The Model Expert

Two distinct forms of knowledge are used by the model expert. Together they distinguish between possible and impossible configurations of features; confidence ratings are then provided for acceptable configurations.

1   Statistical knowledge from a large number of faces, extracted by hand during earlier work [14], forms an invaluable kernel to the model expert. These take the form of feature-point frames and feature frames.

Feature-point frames each represent one of the 37 target points. Our collection of sample images provides a mean and variance for the points' location relative to a fixed point. A predicted location based on the current context is calculated by the model expert, and individual
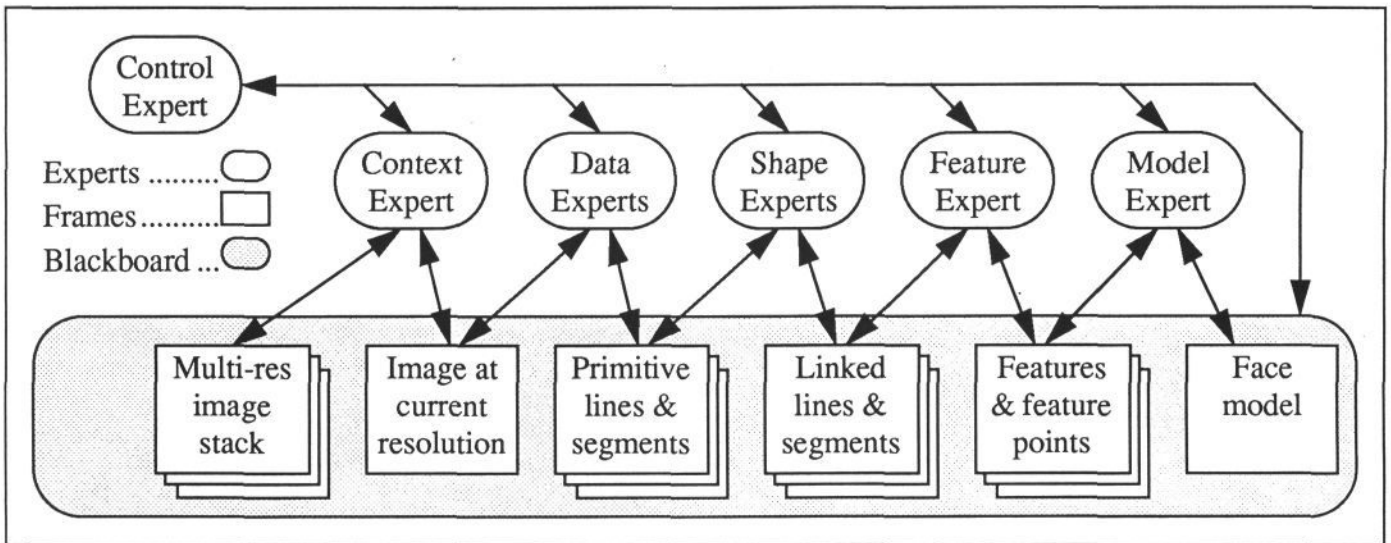
Figure 3: **Schematic view of the blackboard system.**

feature experts provide a third location, based on their examination of the image.

Feature frames serve to group together the individual points that make up a single feature. Their main purpose is to provide a level of indirection between the feature experts and the model expert. Knowledge about the relationship of points within a feature is confined to feature experts, as this allows more flexibility in the representations used.

2 Geometric knowledge supplements our statistical knowledge when the distribution of individual features provides inadequate constraints. Thus statistically feasible configurations, such as a high chin position and a low mouth positions, each individually acceptable, are jointly rejected if the result is a mouth located outside the face outline.

The geometric knowledge takes the form of constraint frames. These identify two or more features to which a constraint procedure is applied. Constraints that apply to only one feature are embedded within the appropriate feature expert. It is sometimes possible to have several constraints that may apply at different times, but only one at any given time. Such is often the case where hair is involved. There may be hair all round the face, only on the top and sides, only on the sides, or not at all. These lead to alternative mutually exclusive constraints.

Evaluating all possible combinations of constraints to determine a consistent set would lead to a combinatorial explosion. This does not happen in practice as the model starts in a consistent state and changes incrementally as features are refined. When evaluating the model following a feature refinement, the model expert need only examine constraint frames involving the refined feature, and reject the refinement if a consistent set of constraints

can not be found.

Basing its reasoning on the three positions in the feature-point frame the model expert selects candidate features and proffers them for refinement. The model expert can be designed to use a number of strategies. It could, for example, always add the features to the list in the same order, irrespective of the values of their positions. This could be done to force a serial processing order as performed by Sakai et al [12, 13]. Other possibilities include selecting features with the minimum estimated variance, hence attempting to locate even more accurately, features which are already believed to be well located, or selecting features with highest uncertainty in their position, hence attempting to locate the least well defined features first.

In practice, none of these *pure* methods works well. The model expert attempts to take all these factors into consideration when selecting features for refinement. There is a built in prejudice for the order or selection, based on the relative importance of the features in human recognition of faces determined by [11]. This is combined with knowledge of the past performance of the feature experts to select features which are most likely to be successfully refined *and* contribute to the developing solution.

Context Expert

The context expert is usually one of the first experts to examine a new input image. It uses one of a number of methods to determine the location and scale of the subject. Based on our knowledge of the input images, an initial estimate may be simply the center of the image, almost filling the image vertically. If this assumption is not confirmed other methods can be used.

Simple thresholding may be sufficient by itself to pro-

403

vide a context; and whilst the eyes are often difficult to locate because of the high frequency detail present, two horizontally adjacent patches of such detail at a suitable separation may represent a pair of eyes, and hence provide a context. Yet another technique is to utilise a robust feature expert, such as the outline expert described below.

If more than one context is obtained, the context expert determines a ranking for each, and selects the highest ranked context as current. Subsequent processing may discredit the current context, in which case the context expert reconsiders the remaining untried contexts and selects one as the new current context.

Feature Experts

The model expert has only configurational knowledge of facial features; it considers the face essentially as a line drawing (figures 4 and 5.) In contrast, feature experts interpret the output from the lower level shape experts, and provide the link with the grey scale image. They have knowledge of how a feature may appear at different resolutions, and can usually employ a number of different methods to achieve their goal.

The model expert identifies candidate regions; these prime the feature experts which in turn consult shape experts, requesting data at specific image resolutions. This allows feature experts both to hunt at low resolution, and yet request very specific data at resolutions up to that of the input image.

Shape and Data Experts

At the lowest level, we have data experts which work directly on the original image, and often correspond to traditional image processing operators. In particular, edge information is provided by a Sobel expert; the Canny [4] edge detector is the basis of another, while a split and merge segmenter [8] provides other shape information. These are implemented in C. Many options have been provided on the image processing operators, giving the data experts a large degree of flexibility. In particular, operations may be applied to particular areas of interest rather than the entire image, resulting in considerable time savings.

Low level image processing techniques of this type usually fail to produce the result expected intuitively, with false identifications complicating the interpretation phase. We thus have another class of expert, the shape expert, which refines the output of data experts. Nazif and Levine [10] describe a rule-based expert system to segment an image; the resulting segmentation is then passed on for interpretation: in our terminology, this would be a shape expert. Since we already know what the image should contain, we extend their idea; although

our shape experts are themselves domain-independent, the tasks they perform, such as, "identify lines in a particular region which curve slowly", are driven by feature experts using high level domain-specific knowledge to guide and interpret the early processing.

Example - The Outline Expert

Although a carefully chosen threshold can sometimes find a head outline provided the background is relatively homogeneous, there is more difficulty in assigning confidence to the resulting shape. The more elaborate method described here is certain to return a head shape, and has a natural confidence level associated with it. The initial data are taken from a Canny edge detector. The data expert selects a suitably fine associated blur, and appropriate thresholds for the hysteresis step. The resulting edgels are passed to a shape expert which links adjacent edgels and determines the longest lines found in the image; perhaps all those longer than one eighth of the height of the image. We have found this to be a useful higher level description of the image; the significant data reduction allowing for a more intensive reasoning process with those data that remain.

The list of long edges is then placed on the blackboard as a higher level primitive, on which other experts can work. An example is the template matcher; given a template defined at a particular resolution, the matcher returns a line which is close to the template, and subject to this, has the highest proportion of its points taken from those on our long edges. To do this at varying scales, each of the long edges is coarsened to produce edges at coarser resolutions, thus giving a multi-resolution form of the output of the "long edge" shape expert. The effect of this is to perform something close to gap spanning on the original Canny output; we prefer this view of the data to that obtained by using a coarser scaled Canny filter on the original image.

Our outline expert starts with an initial estimate of the location of the face. This can be a crude one passed from the model expert, such as is shown in figures 4 and 5, or it can be a more accurate version found by a different expert described below. The knowledge in the outline expert is that outlines can often be built by successive refinement. A coarse version of our initial template, say at a scale of 16×16 pixels is passed to the matcher, which returns an improved location and confidence. This is then scaled up naively to a 32×32 version and in turn serves as a template, or initial approximation to the matcher. Providing our confidence in the result does not decline significantly the process is repeated until full scale is reached, giving the final outline.

If the above process does not give good results, more elaborate means are adopted to obtain the initial approximation. One such method is to use an "average" head outline as the initial template, which is then deformed

probabilistically, based on the known statistical properties of head outlines. Such methods have been described, initially for hands [5]; subsequently work has been done for eyes and mouths [18] and on outlines themselves [3]. Each of the resulting shapes is guaranteed to be appropriately shaped; a choice between them is made based on maximizing the total fitness of the shape. Because we wish to calculate frequently, and require to reward appropriately outlines which are "nearly" on an edge, a fast "edge interest" [16] operator is used at present. The resulting optimization problem - certainly non-convex - is tackled using simulated annealing. Even in this case, our experience has been that multiresolution methods are needed, requiring convergence at a given scale before working at a finer scale. Subject to this, convergence usually occurs in a few minutes, working in C on a Sun 3/50.

We have described this "approximate head shape" feature expert separately for two reasons; it may be unnecessarily expensive to use it - a decision taken by the control expert - and the resulting outline still needs refining by the outline expert to improve on our necessarily weak definition of "convergence".

SYSTEM OPERATION

Once an image is presented to the system, the approximate location of a face is determined by the context expert. Based on the results of earlier runs and the current configuration of features, the model expert generates an ordered list of candidate features for refinement.

The appropriate feature expert is consulted for the first feature on the list; the feature expert attempts to either improve the features' estimated location or reduce the uncertainty. The methods used by the expert depend on the resolution currently being used. Shape experts are consulted to obtain specific details about the area of interest. If no evidence can be found to support the existence of the feature at the estimated location, or the position can not be refined, the expert fails and the feature is removed from the list. The process is repeated for the remainder of the list.

If a feature expert refines the position of a feature, the model expert is once more consulted to evaluate the effect of the refinement. The refinement of one feature may affect the confidence and estimated location of many others. The accurate location of typically 3 or 4 points, is sufficient to tightly constrain the location of the others.

If no further refinements are made, a higher resolution image is selected and the refinement process recommences. Should enough feature experts fail to locate their respective features, the context expert is reconsulted, and the process repeated in a new context.

The results of each stage of processing are stored and used to influence subsequent decisions. In particular,
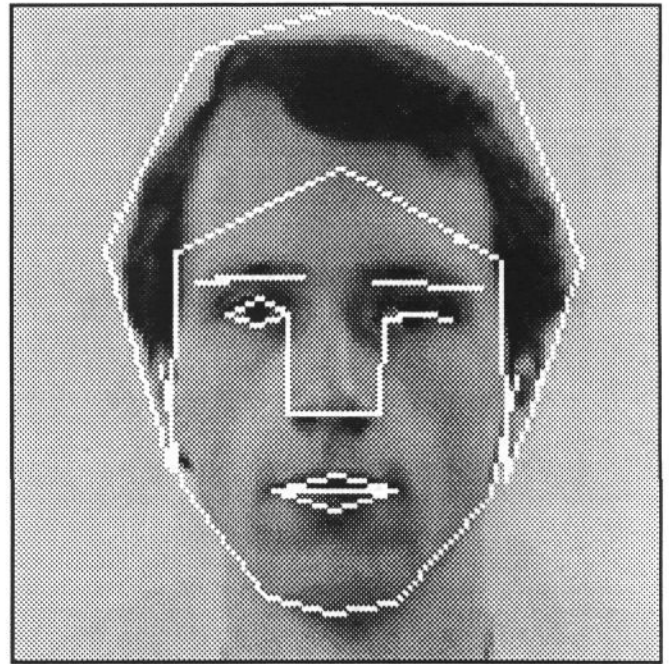

Figure 4: **A sample image with 37 point frame overlayed.**

the order in which features are proffered by the model expert is based largely on the potential for refining its location, and the methods used by feature experts are selected according to previous success rates.

INITIAL RESULTS

The system at present locates the general position of a face, and attempts to refine the mouth[1], the eyes and the eyebrows. A limited number of methods are currently available to these feature experts; the addition of further methods will improve their success rate. Once these initial features are located, the position of the remaining features can be accurately predicted, and as further feature experts are added their locations may also be refined.

Figure 4 and figure 5 show the complete model overlayed on two sample images, based purely on the location of the eyes and mouth. Close inspection of these images reveals that the position of the refined features is within a pixel of the location usually selected by humans; the accuracy of the other features reflects the validity of the model experts knowledge. The hair line in particular is not a good match; this is understandable considering that the hair is one of the most variable and readily varied aspects of a face.

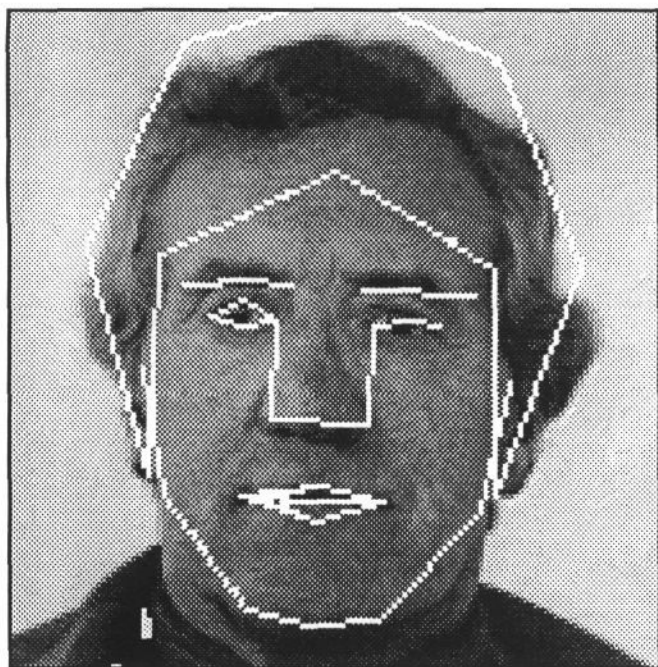[1] Figure 2 shows the right hand corner of the subjects mouth.

405

Figure 5: **Another sample image with 37 point frame overlayed.**

References

[1] **A.J.Goldstein, L.D.Harmon, and A.B.Lesk.** "Man-machine interaction in human face identification". *Bell System Technical Journal*, 51(2):399–427, February 1972.

[2] **Richard Baldock and Simon Towers.** "First steps towards a blackboard controlled system for matching image and model in the presence of noise and distortion". In Josef Kittler, editor, *Pattern Recognition, 4th International Conference*, pages 429–438, Berlin Heidelberg, 1988. Springer-Verlag.

[3] **Alan Bennett.** "Finding Feature Outlines". Technical report, Department of Mathematical Sciences, University of Aberdeen, Aberdeen, Scotland, 1990.

[4] **John Canny.** "A computational approach to edge detection". *IEEE: Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[5] **Y. Chow, U. Grenander, and D.M. Keenan.** "HANDS. a pattern theoretic study of biological shapes". Technical report, Division of Applied Mathematics, Brown University, Providence, RI, 1988.

[6] **I. Craw, H.D. Ellis, and J.R. Lishman.** "Automatic extraction of face-features". *Pattern Recognition Letters*, 5(2):183–187, 1987.

[7] **Robert Engelmore and Tony Morgan.** *Blackboard Systems.* Addison Wesley, Wokingham, England, 1988.

[8] **Steven Horowvitz and Theodosios Pavlidis.** "Picture segmentation by a tree traversal algorithm". *Journal of the ACM*, 23(2):368–388, April 1976.

[9] **Takeo Kanade.** *Computer Recognition of Human Faces*, volume 47 of *Interdisciplinary Systems Research*. Birkhäuser, Basel,Stuttgart, 1977.

[10] **Ahmed M. Nazif and Martin D. Levine.** "Low level image segmentation: An expert system". *IEEE: Transactions on Pattern Analysis and Machine Intelligence*, 6(5):555–577, 1984.

[11] **G. Rhodes.** "Looking at faces: first-order and second-order features as determinants of facial appearance". *Perception*, 17:43–63, 1988.

[12] **T. Sakai, M. Nagao, and S. Fujibayashi.** "Line extraction and pattern detection in a photograph". *Pattern Recognition*, 1:233–248, 1969.

[13] **T. Sakai, M. Nagao, and T. Kanade.** "Computer analysis and classification of photographs of human faces". In *First USA-JAPAN Computer Conference*, pages 55–62, 1972.

[14] **J. W. Shepherd.** "Frame — Coding evaluation 1". Department of Psychology, University of Aberdeen, Progress report of Home Office work.

[15] **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.

[16] **Andrew C. Sleigh.** "The extraction of boundaries using local measures driven by rules". *Pattern Recognition Letters*, 4(4):247–258, September 1986.

[17] **Simon Towers.** "Frames as data structures for SBS". MRC internal paper, 1987. Pattern Recognition and Automation Section, MRC Clinical and Population Cytogenetics Unit, Western General Hospital, Edinburgh.

[18] **A. Yuille, D. Cohen, and P. Hallinan.** "Facial Feature Extraction by Deformable Templates". Technical Report CICS-P-124, Center for Intelligent Control Systems, Cambridge, MA 02139, March 1988.