Sakai 8x

COMPUTER ANALYSIS AND CLASSIFICATION OF PHOTOGRAPHS OF HUMAN FACES

Toshiyuki Sakai, Makoto Nagao, and Takeo Kanade

(Kyoto University, Kycto, Japan)

I. INTRODUCTION

In picture processing and scene analysis by computer, compared with character recognition, we are confronted with much more complicated pattern variations and noise behavior. Therefore it is necessary that the processing should take advantage of structural or contextual information about object patterns in order to achieve good results.

In this paper, a context-controlled picture processing scheme is first presented. After that follows a description of a program based on that scheme to extract feature points from a photograph of the human face, including the eyes, nose, mouth, and chin. The processing results of nearly 800 photographs by the program are also demonstrated.

Human faces are very interesting as objects of automatic picture processing for several reasons:
(1) They are not artificial, and not as simple as cubes or pyramids.
(2) A face has many components constituting substructures: eyes, nose, mouth, chin and so on, which need exact analysis.
(3) Such components are distributed within a certain permissible range in the face, whose mutual relations should be correctly grasped for recognition.
(4) The variety of human faces is as large as the human family.

Research on human-face identification has been done by several people. Bledsoe[1] developed a man-machine system, in which a computer classifies human faces based on the fiducial marks drawn by humans on the projected photographs. Harmon and his group[2] recently constructed an interactive system for face identification that takes advantage both of the human's superiority in describing face features and of the machine's superiority in making decisions based on accurate statistics. Kaya and Kobayashi[3] have discussed, in information-theoretic terms, the number of human faces that can be distinguished by using nine geometric parameters. Studies that deal with faces as picture processing are found in [4] and [5]. Sakai, Nagao and Fujibayashi[4] have written a program to detect faces in a photograph by the use of templates for the mouth, nose, eyes and contour of the head. Kelly[5] tried to identify about a dozen people by measuring a set of parameters out of a full-length photograph of a man.

In brief, our method is as follows; A binary-valued, line-like picture is first obtained out of a gray-level picture of a face by applying a Laplacian operator. Then, the positions of eyes, nose, mouth and chin, and chin contour are detected. The computer program employs the "detection-after-prediction" scheme like Kelly's. The distinguishing feature of our program is that subroutines divided into blocks, each for detecting a part of the face, are combined into action not only by making use of the contextual information of a face, but also by employing a feedback mechanism for corrections and retrials.

II. PATTERN STRUCTURE AND PROCESSING METHOD

The pattern-matching method, we know, is not very effective for recognizing patterns with large variations. In extracting features from a picture, two-dimensional Fourier transforms and histograms of gray levels may be useful, for example, in distinguishing corn fields from woods in aerial photographs, but in cases where the structure of the pattern is of more concern, they do not work satisfactorily. Therefore, it becomes essential to provide the computer with the structural information of patterns. Noticing the analogy between the structure of patterns and the syntax of languages, Kirsch[6], Narasimhan[7], Ledley[8], and others took the so-called linguistic approach. As shown in Fig. 1, the input pattern is first represented in a string of primitive elements which are extracted in a rather microscopic way. Then a syntactic analysis is performed on the string in order to decide whether the given pattern is in the particular class or not. This approach could be successfully applied to line-like patterns, such as handwritten characters, tracks of bubble chamber photographs, and chromosome pictures, achieving more flexibility than the template-matching method or the method based on the statistical decision theory.

Some difficulties will be encountered if one tries to take this linguistic approach in processing photographs of human faces. If small line segments, arcs and the like are selected as the primitives, then the resultant rules for composing a face will be intolerably complicated and awkward. Moreover, a lot of noisy line segments unrelated to the face structure, perhaps from wrinkles and hair style, will be annoying. Thus it will be more convenient to consider a face to be composed of larger components such as eyes, nose, mouth and so on, having the relatively macro-positional relations. Remember that the eyes, for instance, are recognized as such only in the context of a face; otherwise many components would be found in the picture to have the shape of the eye. Therefore it is necessary to search for the component with the shape of the eye

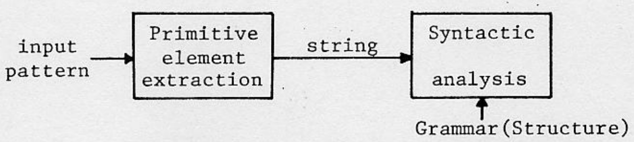


Fig. 1. Introduction of structure in the linguistic approach. This scheme is not satisfactory for complex pictures.

in the estimated region in which the eye must exist. That is, the detection of the primitives or parts should not be separated from the structural-analysis stage as in Fig. 1, but the former needs to be under control of the latter.

The scheme of Fig. 2 is better for this purpose: first the region in which a particular part (e.g.,eye) is predicted to exist is determined by taking full advantage of *a priori* knowledge of the object patterns and of the global but sometimes not so precise information extracted from the input picture. Then the component with specific properties is detected in it. As a result, the information about the picture under processing is reinforced and this can be used, in turn, for determining where to look for the next part. Iteration of these processes leads to the acquisition of information enough to describe the given picture.

Procedures of detecting parts are divided into subroutines and they operate on the estimated search area and under given constraints. For rough and tentative estimation of search areas, procedures insensitive to noise are used which might not be so accurate, but once the search areas for specific components are given, accurate procedures are employed which determine the exact properties of the components. If the program fails to detect the part satisfying the specified conditions in the predicted area, it diagnoses the cause, and may modify the program parameters, or may go back to the former steps. This kind of feedback mechanism makes processing very flexible.

This scheme resembles the top-down goal-oriented syntax analyzer of Shaw[9] in that it employs "detection-after-prediction" process. In our scheme, however, (1) *a priori* structural information about patterns is embedded in the procedures, therefore, (2) the next part to be looked for and its search area are determined by taking into account a broader context than the local head-tail connection in Shaw's analyzer,

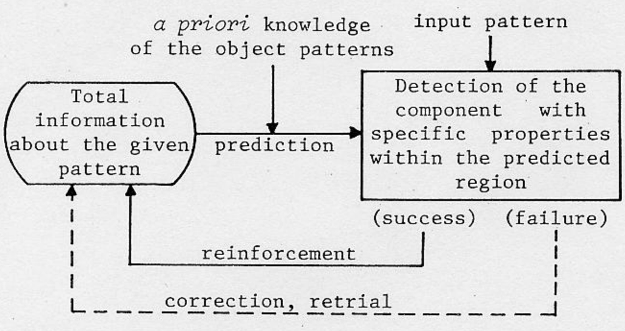


Fig. 2. Context-controlled picture-processing scheme.

and (3) modification of the program parameters and backtracking can be performed more flexibly. In contrast to Shaw's approach it may be a more problem-dependent scheme.

Based on the idea of Fig. 2, a program was written for the analysis of human-face photographs, the details of which will be described in the succeeding sections.

III. INPUT OF HUMAN-FACE PHOTOGRAPHS AND LINE EXTRACTION

What we are going to deal with are photographs of one full face with no glasses or beard. We assume that the face in a photograph may have tilt, forward inclination, or backward bent to certain degrees, but is not turned to the side.

Fig. 3 is the block diagram showing the flow of processing. A photograph (or a real view) of the face is converted to an array of 140 x 208 picture elements, each having a gray level of 5 bits (i.e., 32 levels),by means of ITV camera. Fig. 4(a) is a printout of a digitized gray-level picture. The binary picture of Fig. 4(b) is obtained by applying a Laplacian operator (i.e. two-dimensional secondary differentiation) and thresholding the result at a proper level. This binary picture represents contour portions where brightness changes significantly. It preserves the structural information of the original gray-level picture and is far easier to deal with because of binary values. The analysis which follows is done on this binary picture. We used the 9 x 9 Laplacian operator shown in Fig. 4(c). This operator has been demonstrated in [10] to work very successfully for line extraction of human-face photographs because it has the effect of eliminating noise by averaging a large area.

IV. FACE-FEATURE EXTRACTION

In Fig. 5 is shown the logical connection of subroutines in the analysis program of human-face photographs. In general, the analysis steps proceed from easy to difficult, and from grasping approximate, global information to detecting accurate positions of eyes, nose, mouth and so on. Each subroutine shown by the square box tries to detect a component (e.g., eye) satisfying specific conditions (e.g., location, size and shape) in the predicted region. According to whether it succeeds or fails in the detection, the subroutine to be executed next is determined and various program parameters are modified. A typical order of processing steps is from ⓐ to ⓗ as in Fig. 6.

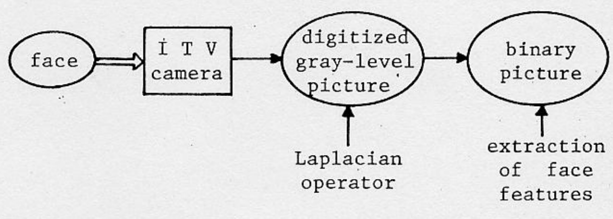


Fig. 3. Block diagram of processing of human-face photographs.

(a) (b)

			1	1	1			
			1	1	1			
			1	1	1			
1	1	1	-4	-4	-4	1	1	1
1	1	1	-4	-4	-4	1	1	1
1	1	1	-4	-4	-4	1	1	1
			1	1	1			
			1	1	1			
			1	1	1			

(c)

Fig. 4. (a) Printout of a digitized gray-level picture.
(b) Binary picture that represents the contour portions.
(c) The 9 x 9 Laplacian operator used for line extraction.

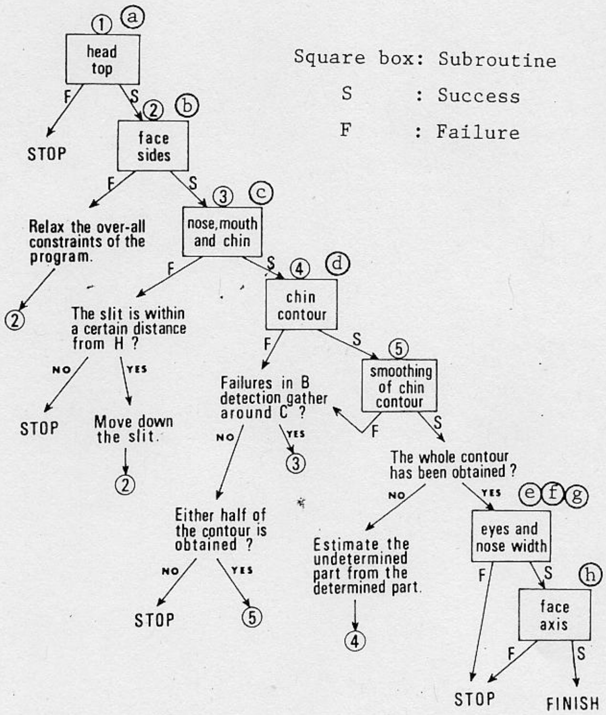


Fig. 5. Logical connection of subroutines

A fundamental, useful technique of picture processing used throughout the program is an "integral projection". As shown in Fig. 7, a slit of proper width and length is placed in a picture. A histogram is obtained along the length of the slit by counting the number of elements " * " in the direction of the width. This will be called an integral projection (curve) of the slit. If the slit is applied within a suitable area with the proper direction and width, the integral projection tells reliably the position of a component in the picture even in the presence of noise.

Now, the analysis steps will be described.

1) Top of Head

A horizontal slit is moved down from the top of the picture. The first position with sufficient output is presumed as the head top H. This position is used only for setting the starting point of slit application in the next step and therefore it need not be so precisely determined.

2) Sides of Face at Cheeks

Starting from the point of a certain distance below H, a horizontal slit of width h is applied successively, shifting downward with $h/2$ overlap as in Fig. 8. When the slit crosses the cheeks, its integral projection displays characteristic patterns such as the lowest one of the three in Fig. 8(a): two long clearances and one or two peaks sandwiched by them correspond to the cheeks and the nose, respectively. Thus the left and right sides of the face at the height of cheeks are obtained. They are indicated by L and R in Fig. 8. Since we need to know these positions only approximately, the width h of

Fig. 6. Typical order of the analysis steps:

- ⓐ top of head
- ⓑ cheeks and sides of face
- ⓒ nose, mouth, and chin
- ⓓ chin contour
- ⓔ face-side lines
- ⓕ nose lines
- ⓖ eyes
- ⓗ face axis

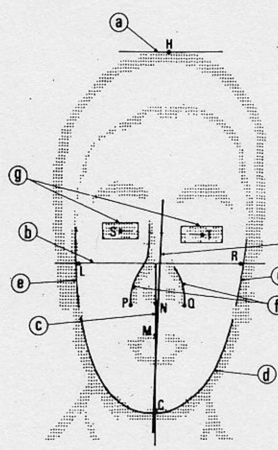
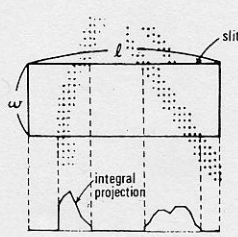


Fig. 7. Integral projection of a slit. The curve is obtained along the length of the slit by counting the number of elements " * " in the direction of the width.



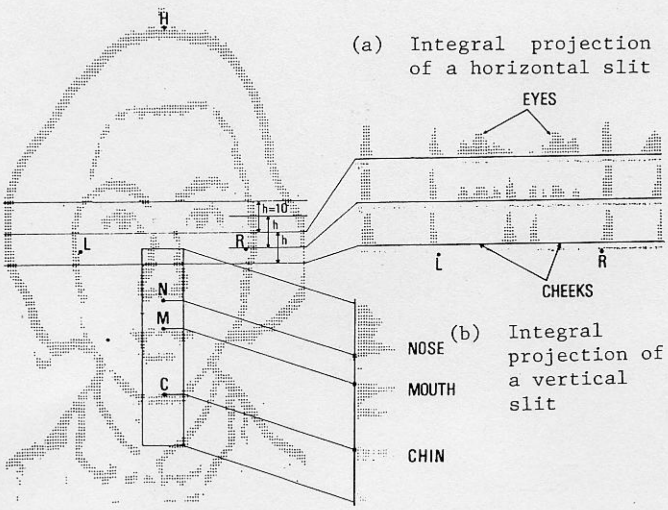


Fig. 8. Detection of the face sides, nose, mouth, and chin by the application of slits.

the slit is set rather large so that it can pick up thin line portions. Here we set h = 10.

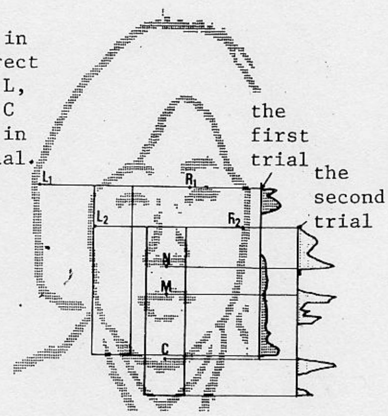
Erroneous detection of L and R in this step will cause the succeeding step 3) or 4) to fail in detecting the expected components, and thus the analysis will come back to this step again to obtain the correct positions of L and R.

3) Vertical Positions of Nose, Mouth and Chin

A vertical slit of width LR/4 is placed at the center of L and R as shown in Fig. 8. A nose, a mouth and a chin are probably in it if L and R are approximately correct and if the face does not tilt or turn a great deal. It is observed that in the integral projection curve of the slit the peaks appear which correspond to the nose, mouth and chin. Fig. 8(b) is one of the typical types of the curve. The integral projection curve is then coded into a compact form by using the symbols for the types of the peak and its length. On the other hand, several types of the standard integral projection are also coded in the same manner and stored in the variable-length table. At present, eleven standard types of the coded integral projection are prepared. The table can be extended simply by adding the new types. The integral projection of the input picture is compared with standard types. If it matches with only one type, the vertical positions of the lower end of the nose N, the upper lip M and the point of the chin C are obtained. In case there is no match, the matching criteria of the standard types are slightly relaxed, or the vertical slit is shifted a little. Matching is tried again. If there is no match even then, this step is a failure.

In the case of failure, considering that L and R were located incorrectly in the former step, the program goes back to the step 2) and tries to find out another possibility for L and R. Fig. 9 is an example in which the correct results were obtained in the second trial. We could correctly determine these positions in this way for many pictures for which the analysis was unsuccessful in the first trial. It is

Fig. 9. An example in which the correct positions of L, R, N, M, and C are obtained in the second trial.



notable that this trial-and-error process with the feedback mechanism allows each analysis step to be simple. The once-and-for-all process would have required that each step be highly elaborate and reliable, so as not to deliver erroneous results to the next step.

Pattern matching applied on the restricted region in a picture has a positive meaning in contrast to that applied on the whole picture; a successful match means that what was expected exists in the predicted region. It increases the reliability of the analysis results so far obtained and used for prediction as well as adding the new result. On the other hand, failure in matching may require that some of the program parameters be modified or that a feedback procedure be activated for corrections and retrials.

4) Chin Contour and Smoothing

In obtaining the chin contour, a method like tracing it from one of its ends would probably lead in the wrong direction because of the existence of line splits and extra lines, even if elaborate devices such as line extension are employed. Such a method is substantially based on the local decision, i.e.,line connectivity. Since we have already known, at least approximately, the face sides L and R, the nose N, the upper lip M and the chin C, we can limit the search

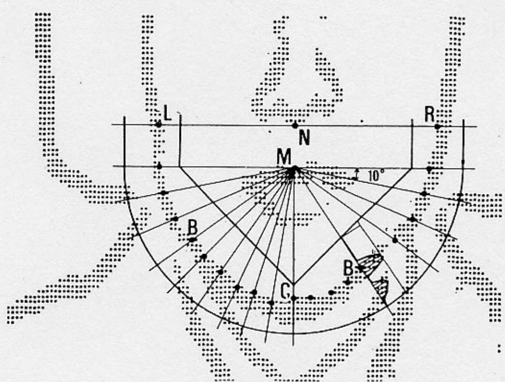
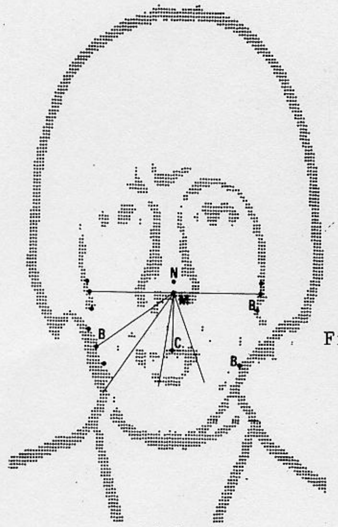
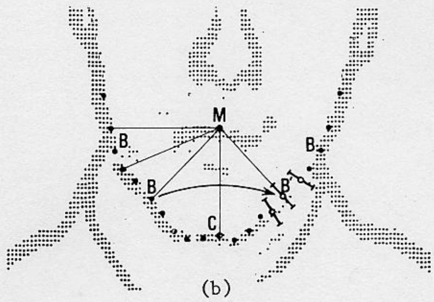


Fig. 10. Extraction of the chin contour. The search area is established and a slit is placed along each radiation line.



(a)



(b)

Fig. 11. Recovery from the failures in the detection of chin contour:

(a) the case in which N, M and C were misidentified;

(b) the case in which a part of the chin contour was not obtained. The B' is guessed by the symmetry.

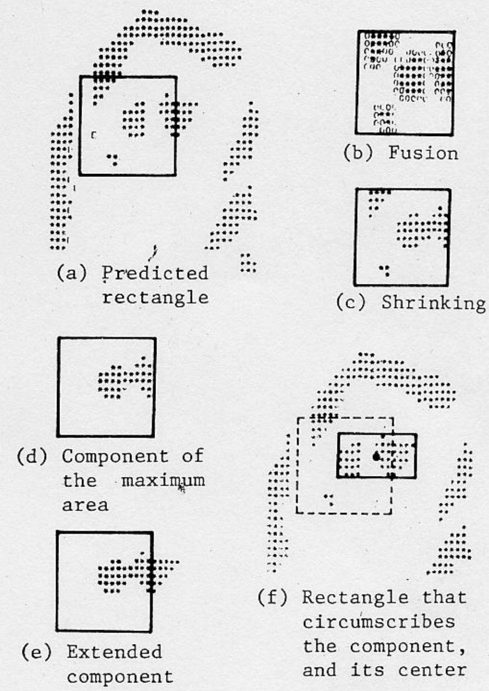


Fig. 13. Determination of the eye center.

area for the chin contour. As shown in Fig. 10, the search area is determined by L, R, C, M and N. Nineteen radiation lines are drawn from M downward every 10 degrees. These lines are expected to cross the chin contour in the predicted area at nearly right angles, if the positions of L, R, C, M and N are correct. A slit is placed along each radiation line and its integral projection is examined. A contour point B is detected as the peak nearest to M within the specified area. The sequence of B's thus obtained for nineteen radiation lines is smoothed. Limit imposed on the search area and the smoothing enable this step to avoid mistaking long wrinkles or neck lines for a part of the chin contour.

In the B point detection, B is not determined on the radiation line whose incidental integral projection contains no conspicuous peak within the search area. In case B's cannot be determined on three consecutive radiations, the contour detection is judged unsuccessful. The program then returns to the main routine. The cause is investigated and an appropriate measure is taken for each of the following three cases:

(1) N, M and C were misidentified in the former step 3) as upward shifted positions as in Fig. 11(a). Usually failures in B detection occur around C in a symmetrical manner. When this is found, the feedback mechanism forces the program to go back to the step 3) to re-examine the positions of N, M and C.

(2) There exists a short break in the contour, or the real contour line partially sticks out of the search area. In this case, either half of the contour usually is obtained correctly. Thus, for the radiation line whose B was not obtained, a prediction point B' is determined there by utilizing the symmetrical property with respect to the vertical axis as shown in Fig.11(b). The same routine for chin-contour detection is again executed with the narrower search region around B' and with the lower threshold for peak detection. By this process it might be possible to detect the thin portion of the contour which was undetectable in the first trial.

(3) The other cases include one in which there was not sufficient contrast around the chin in the original gray-level picture. In these cases, the remainder of the analysis is resigned.

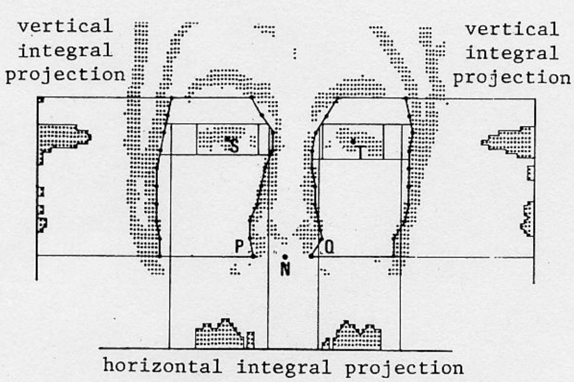


Fig. 12. Cheek areas and rectangles which contain the eyes.

5) Eye Positions, Nose Width, and Face Axis

As is shown in Fig. 12, by successive application of horizontal slits starting from N, we can trace the nose upward as well as the left and right face sides up to approximately the eye position. Both ends of the nose, P and Q, are determined tentatively. Then the left and right cheek areas are determined. Examination of the vertical and horizontal integral projections of these areas gives the rectangles containing the eyes.

As in the case of Fig. 13(a), the rectangle may contain a part of the eyebrow, the nose or the face-side line. Even a part of the eye may lie outside of it.

The following operations determine the position of the eye precisely: first, the operations of fusion and shrinking[11] are executed in the rectangle cut out of the picture to eliminate gaps and hollows which often appear in the eye area (Fig. 13(b),(c)). The connected component of the maximum area is then identified in the rectangle (Fig. 13(d)). The parts which are outside the rectangle, but which are connected to this component, are joined, and the fusion and shrinking are performed again on this extended component (Fig. 13(e)). The position, the size, and the shape of the component are examined. If they satisfy certain conditions, the eye center is determined as the center of the rectangle which circumscribes the component (Fig. 13(f)).

In this way, the left and right eye centers, S and T, are obtained. By applying a slit on N parallel to the direction ST, the nose ends, P and Q, are redetermined. The first approximation of the face axis is the line through X and Y, which are the center point of S and T, and that of P and Q, respectively. The orientation of the face is corrected by the measurements not only of the line XY, but also of the face-side points obtained by the application of several slits vertical to XY.

V. PERFORMANCE OF THE PROGRAM AND FACE PARAMETERS

About 800 photographs of human faces have been processed by the program described above. They include faces of young and old, males and females, with glasses and hats, and faces with turn, tilt or inclination. It is of great interest to see how the program works on them. The results are summarized in Table 1. For photographs fulfilling the constraints described in the first part of section III, 552 photographs out of 607 were successfully analyzed, giving all of the face-feature points correctly. For about 100 of them the analysis failed halfway once or twice, but the feedback mechanism including diagnosis, correction, and retrial, made recovery possible.

As is seen in Table 1, the program works fairly well also on photographs which do not satisfy the presumed constraints. In the case of faces with glasses, the program fails or gives erroneous eye positions, but this is only natural in the sense that the program can not find a component with the shape of the eye in the predicted region. All the other results, however, are usually correct.

A few examples of the results of analysis are given in Fig. 14. We can observe how the functions and devices incorporated work in extracting feature points; limiting the search area for chin contour, rough prediction of eye positions, and fusion and shrinking in detecting eye centers.

A set of parameters which seem to characterize faces are defined as illustrated in Fig. 15. These parameters are calculated for all the faces for the purpose of classification of faces. Faces that have extreme values of parameters are shown in Fig. 16 to demonstrate some correspondences between those parameters and human impressions of singularity. The i-th component y_i of the vector below each face shows how far the parameter x_i is from the mean $\hat{x}_i$, normalized by the standard deviation σ_i, i.e., $y_i=(x_i-\hat{x}_i)/\sigma_i$. Each pair of the faces (upper and lower) forms positive and negative contrasts in the underlined components. The most conspicuous contrasts are : (a) y_1; oval and elongated faces (b) y_2; wide-set or close-set eyes, (c) y_4; sharp and flat chins, (d) y_5; angular and tapered jaws, and (e) y_7; vertical-slanted and vertical-straight face sides. The research in this direction is now in progress and the usefulness of the parameters will be investigated more thoroughly.

VI. CONCLUSION

In this paper were described the computer analysis of photographs of human faces and some preliminary results toward computer classification of human faces.

The picture-analysis program employed the context-controlled processing scheme described in section II. This scheme, when advanced, leads to the view that a picture-processing program should be a problem-solving program: that is, the processing does not proceed in the predetermined order, but the program decides the best sequence of processing steps depending on individual input pictures, using *a priori* knowledge and the information so far obtained about the given input.

category of faces	number of faces	correct results	error or unrecovered failure	step in which the error or unrecovered failure occurred					
				face sides	nose mouth chin	chin contour	eyes	nose width	face axis
full face with no glasses or beard	607	552	55	5	10	15	18	3	4
full face with glasses	77		77	4	4	2	67		
turning face with no glasses	79	63	16	4	3	3	5		1
face with beard	25		25		10	4		11	

Table 1. Results of the analysis of human-face photographs.

(a) (b) (c) (d)

Fig. 14 Examples of the results of analysis: original gray-level pictures(upper) and processed binary pictures(lower). The dots in the binary pictures are the feature points extracted.

The program selects from a collection of subroutines one that is expected to give the most reliable result. The selected subroutine is executed and the total information about the input increases. In this way, the program reaches the goal of describing the given pattern. In its process, the several aspects of problem-solving techniques will appear inevitably, such as evaluation of intermediate results, subgoal setting, backtracking and so on. The problem of data structure in this method is discussed by Nagao[12].

ACKNOWLEDGEMENTS

The authors wish to thank T. Fujimaki and M. Murao of NEC for gathering the photographic data used in this study, and M. Ohira for his assistance in writing the program.

Fig. 15. Face parameters.

$x_1 = \frac{1}{2}(OE+OF)/OC$

$x_2 =$ ST/EF

$x_3 =$ NC/OC

$x_4 =$ curvature of the bottom of the chin

$x_5 = (\Delta EGC+\Delta FCH)/(S_\ell+S_\gamma)$

$x_6 = \alpha$

$x_7 = \frac{1}{2}(\Theta_\ell+\Theta_\gamma)$

$x_8 = \beta$

S_ℓ, S_γ : hatched areas

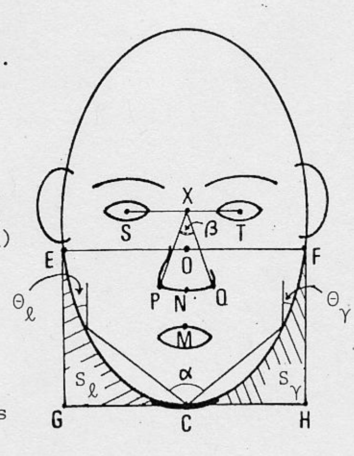


Fig. 16. Faces having extreme values of parameters. The i-th component y_i of the vector below each face shows how far the parameter value x_i is from the mean $\hat{x}_i$, normalized by the standard deviation σ_i, i.e., $y_i = (x_i - \hat{x}_i)/\sigma_i$.

REFERENCES

[1] Bledsoe, W. W., "Man-Machine Facial Recognition", Panoramic Research Inc., Palo Alto, Cal., Rep. PRI:22 Aug., 1966.

[2] Goldstein, A. J., Harmon, L. D. and Lesk, A. B., "Man-Machine Interaction in Human-Face Identification", *Bell Sys. Tech. J.*, pp. 399-427, Feb., 1972.

[3] Kaya, Y. and Kobayashi, K., "A Basic Study on Human Face Recognition", *International Conf. on Frontiers of Pattern Recognition*, Hawaii, 1971.

[4] Sakai, T., Nagao, M. and Fujibayashi, S., "Line Extraction and Pattern Detection in a Photograph", *Pattern Recognition*, 1, p. 233, March, 1969.

[5] Kelly, M. D., "Visual Identification of People by Computer", Stanford Artificial Intelligence Project Memo AI-130, July, 1970.

[6] Kirsch, R. A., "Computer Interpretation of English Text and Picture Patterns", *IEEE Trans. EC* 13, pp. 363-376, Aug., 1964.

[7] Narasimhan, R.,"Labeling Schemata and Syntactic Descriptions of Pictures", *Information and Control*, 7, pp. 151-179, 1964.

[8] Ledley, R. S., "High-Speed Automatic Analysis of Biomedical Pictures", *Science*, Vol. 146, No. 3641, pp. 216-223, Oct., 1964.

[9] Shaw, A. C., "Parsing of Graph-Representable Pictures", *JACM*, Vol. 17, No. 3, pp. 453-481, July, 1970.

[10] Sakai, T., Nagao, M. and Kidode, M., "Processing of Multi-Leveled Pictures by Computer -- The Case of Photographs of Human Face -- ", *Jour. Inst. Elect. Comm. Engrs, Japan, Abstracts*, 54-6, p. 27, June, 1971.

[11] Rosenfeld, A., *Picture Processing by Computer*. Academic Press, New York, 1969, p. 141.

[12] Nagao, M., "Picture Recognition and Data Structure", *IFIP Conf. on Graphic Languages*, Vancouver, May, 1972.