PICTURE PROCESSING SYSTEM BY COMPUTER COMPLEX

AND

RECOGNITION OF HUMAN FACES

Takeo Kanade

Department of

Information Science

Kyoto University

November 1973

PICTURE PROCESSING SYSTEM BY COMPUTER COMPLEX AND RECOGNITION OF HUMAN FACES

TAKEO KANADE

ABSTRACT

The research here described is concerned with development of a picture processing system, and computer analysis and identification of human faces.

The developed system consists of several pictorial-data input/output devices under control of a high-speed minicomputer which is connected to a medium-sized host computer. It has been designed to be as flexible as possible so that it can facilitate the solution of several classes of problems in picture processing. Many examples of usage of the facilities demonstrate a wide and potential applicability of the system.

Pictures of human faces are successfully analyzed by a computer program which extracts face-feature points, such as nose, mouth, eyes and so on. The program was tested with more than 800 photographs. Emphasis is put on the flexible picture analysis scheme with feedback which was first employed in the picture analysis program with remarkable success. The program consists of a collection of rather simple subroutines, each of which works on the specific part of the picture, and elaborate combination of them with backup procedures makes the whole process flexible and adaptive. An experiment on face identification of 20 people was also conducted.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Professor Toshiyuki Sakai for his adequate guidance and encouragement to enable this thesis to be completed.

I am also grateful to Professor Makoto Nagao for having had enlightening discussions with me.

Other people who so kindly extended their assistance and co-operation in my preparation of this thesis are: Dr. Shigeharu Sugita, Dr. Koh-ichi Tabata, Dr. Kenji Ohtani, Shinji Tomita, Masatsugu Kidode, Yuichi Ohta, Hiroyuki Kaji, Hideo Hasui, Takayuki Nagano, and my colleagues in Prof. Sakai's laboratory. In particular, Yuichi Ohta contributed significantly to the technical content of CHAPTER II.

In addition, I should extend my thanks to hundreds of people, the photographs of whose faces supplied indispensable data used in my research.

Constant encouragement and editorial assistance given by Yukiko Kubo are worth mentioning with my thanks.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Picture processing by computer has found its application in various fields. Character recognition has shown the most practical success. Furthermore, the techniques span much more sophisticated applications such as interpretation of biomedical images and X-ray films, measurement of images in nuclear physics, processing of a large volume of pictorial data sent from the satellites, etc. The importance and benefit of automatic picture processing continue to grow.

This thesis is concerned with picture processing of this kind. The contents are, roughly speaking, twofold:

(1) Development of picture processing system by computer complex

The system consists of several pictorial-data input/output devices under control of a high-speed minicomputer which is connected to a medium-sized host computer. It has been designed to be as flexible as possible so that it can facilitate the solution of several classes of problems in picture processing. In addition to the noticeable capabilities of picture processing, the system embodies an interesting form of computer-computer communication.

(2) Pictorial recognition of human faces

Pictures of human faces are successfully analyzed by a computer program which extracts face-feature points, such as nose, mouth, eyes, and so on. The program was tested with more than

800 photographs. The success of the program is due to the employment of a flexible picture analysis scheme with feedback. An experiment on face identification of 20 people was also conducted.

I-1. Picture Analysis and Recognition —— New Aspects

When shown the pictures of the human face of Fig. 1-1, we can immediately tell the positions of the nose, mouth and eyes; and moreover, we can say that both pictures surely portray the same person. Picture analysis and recognition by computer concerns itself with this type of two-dimensional image processing. In this thesis, I selected human-face photographs as objects of processing.

Figure 1-1 Pictures of human face.

Character recognition by machine has been investigated very intensively as one of the most fundamental problems with practical value. Today computer processing of visual images has come to deal with more sophisticated problems: scene analysis in robot projects, biomedical or meteorological image interpretation, recognition of faces or fingerprints, etc.

It might be considered that recognition of faces or other natural scenes would be only a little more complex than characters, but actually, completely new aspects appear.

In the case of printed characters, one can assume the standard pattern. Thus processing usually involves matching with the standard pattern or feature measurements at predetermined points, and decisions about the existence of features are based on the combinational logic.

On the other hand, in processing complex natural images such as faces, the situation is further complicated. There is such a wide variety in the input images that the measurement of features must be adaptive to each individual image and relative to other feature measurements. In addition to these low-level processings —— so to speak "local" processing —— the system needs decision algorithm of higher level —— it should be referred to as "global" recognition — ; it has to interpret the structures of the picture.

The picture analysis program for human-face photographs, developed in this thesis, shows some of these new aspects of advanced picture processing. It employs a new flexible analysis scheme which elaborately combines local processing with global recognition. Backup procedures with feedback and retrial play a very important role in it. Since the variety of pictures to be treated is enormous, fixed analysis algorithms are not very powerful. The process must be flexible and adaptive in the sense that the analysis in a certain portion of a picture is carried out repeatedly with a change of parameters until it gives a satisfactory result. If the analysis fails, its cause is examined in detail to find out the part where the analysis must be performed once more. The flexibility of the human-face analysis program results certainly from these backup procedures.

This is one of the first works that actually implement such a flexible analysis scheme in the picture analysis program. The detail will be described in CHAPTER III.

In CHAPTER IV, computer identification of 20 people by their faces is attempted, using two pictures for each person. A set of facial parameters is calculated for each picture on the basis of feature points located by the picture analysis program. In the identification test, 15 out of 20 people were correctly identified. This result is compared with the case in which a human operator locates the feature points, and it is demonstrated that computer measurement is almost as good as human measurement. The two-stage process employed in the computer feature extraction includes a "distant" feedback that involves both high-level decision and low-level processing.

I-2. Picture Processing System

It should be noted that a picture processing system should be a total system: it needs all of picture input, storage, display, processing capabilities and man-machine interaction. As discussed above, processing capabilities include not only bit-operation to deal with gray-level data, but also symbol manipulation to make high-level decisions.

To fulfill the requirements, a new picture processing system has been designed and constructed: it consists of several pictorial-data I/O devices which are under control of a high-speed minicomputer that is connected to a medium-sized general-purpose computer with filing devices. The system employs system-subsystem organization. The minicomputer and I/O devices form a satellite subsystem named GIRLS (Graphical-data Interpretation and Reconstruction in Local Satellite), which not only functions as a pre-/post-processor but also can carry out some intelligent jobs independently of and/or cooperatively with the host: data input, data transformation, feature measurement, man-machine interaction, film output, etc. Such functions are called into action by an operator's command entered from the teletypewriter or by a program command sent from the host computer. The host computer deals with reduced data, manipulates the major data structure, makes high-level decisions for global recognition, stores data in the file, or prints out the results on the line printer.

The developed system offers a wide variety of facilities for use in solving several classes of problems of picture processing. The design philosophy and functional capabilities of the system will be described in CHAPTER II. Several examples of the potential usage of the system are also presented.

I-3. Related Works

A tremendous number of papers have been published on pictorial pattern recognition or picture processing: excellent survey articles can be found in Nagy[23]. The objective here is not to cover all the aspects of picture processing, but to refer selectively to the works which are closely related to the contents of this thesis. There are two categories:

(1) articles that describe a picture processing system, not a device itself, having some distinguishing features.

(2) articles that deal with human faces as the processing object.

Ledley, Rotolo, Belson, et al.[20] constructed FIDAC (Film Input to Digital Automatic Computer), specifically for processing of bio-medical pictures. Under computer-program control, it has a high-speed scan of less than 0.3 seconds per frame, sampling 700 x 500 points in the seven-level gray mode or 1000 x 800 points in the black-and-white mode. The speed of scanning and the capability of real-time operation make possible the rapid processing of pictures for statistical analysis and screening purposes. The fact that FIDAC is on-line with the computer, with no intermediate magnetic-tape recording, results in extreme flexibility, convenience, and economy of storage of the original data. They also have developed computer-programming systems to facilitate picture analysis and measurement.

Several working systems exist to process bubble chamber photographs produced in high-energy physics. Some of them have very special provisions for scanning.

The PEPR hardware at MIT(Pless[27]) is a CRT scanner that uses a flying *line segment* to locate track elements. The line segment can be oriented at any angle and moved either in the x or y direction. Because the sharpest output occurs when the flying line segment is tangent to the track, both the coordinates and the angle of the track element can be measured very accurately; they are used to analyze bubble chamber events (Bastien and Dunn[3]).

Another interesting manner of scanning is POLLY(Barr[1], Beck[4]). The spot produces parallel short scanning lines which are perpendicular to a track. This permits reliable detection of the track with short breaks.

Pingle and Tenenbaum[26] developed an accommodating edge follower. The program locates objects in a television image and traces their edges. For doing this, it can accommodate the television camera to obtain the most appropriate image for each phase, i.e., maximizing dynamic range during acquisition, and sensitivity during trancing. The capability of an "intelligent" data sampling of this kind is very essential for advanced picture processing.

The digital scanning I/O device at NHK(Fukushima and Yamaguchi[10]) is fully controlled by a small computer. It can digitize film data and record computer output on the film using a single CRT. It is a basically hardware-oriented system where the size and number of picture elements are controlled by a special hardware.

GRAFIX I of Information International Incorp.(Gray[13]) is probably one of the largest-scale and most complete image processing systems which are commercially available. The central processor is a time-sharing PDP-10 with a large core memory and all the peripherals. The system consists of three image-oriented portions: GRAFIX console subsystem which provides programmers and operators with facilities for interaction with the system; a high-precision film-to-digital transducer; and Binary Image Processor (BIP) which is a special-purpose computer, highly powerful for such typical operations in picture processing as smoothing and hole filling.

Many of those capabilities which are judged to be valuable have been implemented in our system, but they are mainly realized in terms of the software of the high-speed minicomputer. In this sense, our system can be described as software-oriented.

Herbst and Will[15] describe the experimental laboratory at the IBM T.J.Watson Research Center, which includes two flying-spot scanners and a TV camera specially interfaced to a process control digital computer, dot-mode and vector display, and analog I/O facilities. Their three design principles (man-machine interaction, direct program control and satellite operation) underlie also the design of our system.

The system-subsystem arrangement similar to the organization we adopted is often employed in computer graphics systems(Foley[9]). DESIGN PAD of IBM (Belady, Blasgen, Evangelisti and Tennison[5]) is a typical example. The host, a time-shared IBM 360/67, is the central analysis machine, whereas the satellite, an IBM 1130 combined with a display unit, is a high-capability user terminal. The satellite communicates with the host via standard telephone service at 2000 bit/sec. This is a low-speed intercomputer communication. The contrast to our system exists in that the host-satellite communication in our system is much more intimate and the high-speed transmission is possible up to 750 kbit/sec by means of the channel-to-channel connection; a large volume of image data can be transferred within a reasonable time delay.

In spite of the fact that it is a matter of everyday practice to recognize faces, machine recognition of human faces has not been attempted so often as recognition of characters or other picture interpretation tasks.

The work begun by Bledsoe(Chan and Bledsoe[7], Bledsoe[6]) is the first important attempt. It is a hybrid man-machine system: a human operator locates the feature points of the face projected on the RAND tablet, and the computer classifies the face on the basis of those fiducial marks. The parameters employed are normalized distances and ratios among such points as eye corners, mouth corners, nose tip, head top, etc.

Kaya and Kobayashi[16] discussed in the information theoretic terms the number of faces that can be identified by this kind of parametrization.

Recently, Harmon and his associates did a systematic research concerned with both the human and the computer's ability of human-face identification(Harmon[14], Goldstein, Harmon and Lesk[11]). After two preliminary experiments they constructed an interactive system for human-face identification(Goldstein, Harmon and Lesk[12]).

A human operator is given a photograph of one member of the population. He describes this target face to the computer using rather subjective "features" like long ears, wide-set eyes, etc. The computer searches through the population to retrieve the "target" on the basis of this subjectively assigned feature judgment. In this way the system takes advantage both of the human's superiority in describing features and of the machine's superiority in making decisions based on accurate knowledge of population statistics.

Those works mentioned above do not deal with human faces as the object of picture processing, and the problem of automatic feature extraction from a picture of the face has been left open for further research.

The first serious work in this direction is perhaps one by Kelly[17], which is most closely related to the work reported here. He uses five measurements from a picture of the entire body such as height, width of shoulders, etc. and five measurements from a close-up of the head such as width of the head, distance between eyes, etc. His program could identify about ten persons whose photographs were taken in a standardized manner. Various methods are used in recognizing each part of the picture: subtraction, edge detection, template matching, and dynamic threshold setting. All of these methods are applied heuristically in a goal-directed manner.

Much more intensive study is reported in this thesis. Only a picture of the face is used. Accurate feature points are located in it, from which the measurements on facial parameters are derived. Though the general flow of processing resembles that of Kelly, an important and clear difference exists in the picture analysis program. The analysis program is not a straight-line program that processes an input picture in a predetermined order, but it employs a flexible picture analysis scheme with feedback. The success of the program has shown that the scheme is very powerful for processing of complex pictures.

Before Kelly's work, Sakai, Nagao and Fujibayashi[31] have written a program for finding faces in photographs by template-matching method. They first produce a binary picture which contains the edges of the input picture. A template for a human face is divided into several subtemplates each of which corresponds to the head line, forehead, eyes, nose or mouth. Those subtemplates are matched with the edge picture. A set of points with high matching output whose mutual spatial relationship is reasonable assure the existence of a face.

Fischler and Elschlager[8] describe a sophisticated method for detecting the global structure by a set of local template matchings.

Suppose that an object (e.g., a face) is composed of several components (e.g., eyes, nose, mouth, etc.) which must satisfy some spatial relationships in order to be recognized as such. A plausible model of this is a set of mass points which are mutually connected by springs: the degree of relationship between two components is represented by "spring constant" of the spring that connects them, and the extent to which the relationship is not satisfied is regarded as "stretching the spring". The evaluation function combines both the degree of local matching of each component and the cost of stretching the springs. Given an input image, the locations of components are determined such that the evaluation function is minimal by using a kind of dynamic programming. They applied this method to the problem of locating feature points in a face.

These two methods have some parallelism in their nature and may be effective to detect an object in a scene; but they appear to be time-consuming, and yield only approximate locations of components. The method employed in this thesis is completely different. The components are detected in rather a top-down manner directed by the model of the face embedded in the program.

CHAPTER II

NEW FLEXIBLE PICTURE PROCESSING SYSTEM BY COMPUTER COMPLEX

II-1. Introduction

In this chapter, the system requirements for picture processing research are first summarized. Based on those considerations, a new picture processing system has been designed and constructed: it consists of a flying-spot scanner, a Graf/pen(coordinate-input device of tablet type) and a storage-type CRT, which are under control of a high-speed minicomputer that is connected to a medium-sized general-purpose computer with filing devices. The system employs system-subsystem organization; the minicomputer and input/output devices form a satellite subsystem called GIRLS (Graphical-data Interpretation and Reconstruction in Local Satellite), which not only functions as a pre-/post-processor but also can carry out some intelligent jobs independently of and/or co-operatively with the host machine. The whole system has several features in both hardware and software, so that it can be used in a flexible manner for the solution of several classes of picture processing problems.

This chapter provides only a general descriptive presentation of the system. The detailed descriptions of the developed hardware and software are found in APPENDIX A. In the last section of this chapter, an on-line analysis of human-face photographs is demonstrated as an example of a potential use of this system.

II-2. Basic Design

II-2-1. System Requirements for Advanced Picture Processing Research

The following are some of the basic requirements to be considered in designing a new picture processing system.

(1) flexible input of image data under various program-controllable conditions

(2) high-speed, on-line, real-time control of devices

(3) intimate man-machine interaction

(4) file, list-processing, and problem-solving techniques should be available: point-by-point operations such as thinning, thresholding and the like are not all of the picture processing. Higher-order decision-making is essential for advanced picture processing.

(5) As a research instrument, the system should be soft: system modification and expansion seeking for better performance and capability are to be made easily.

Let us discuss these points in some detail. As the computer picture processing comes to cover broader applications, necessary accuracy and resolution become higher and the amount of data bits to represent one frame of a picture is enormous. For instance, a complete scan of a frame of 1000 x 1000 point array with 5 $\sim$ 8 bit gray levels per point produces 5 $\sim$ 8 x 10^6 bits. Furthermore, data sampling under various settable conditions such as resolution, contrast, offset and so on, or color-image sampling will give rise to an unmanageable amount of stored data. Sometimes to store image data on magnetic tapes is not feasible: consider the case of a microscope image which needs three-dimensional resolution; the data may amount to 10^{11} bits[34]; it would fill a thousand reels of 2400 ft magnetic tape!

This leads to the view that image data are stored most compactly in the original films or pictures, and that it is more advantageous to enjoy "an intelligent scan"; when the processing is just being done, only relevant, interesting portions of the picture are scanned under the necessary and sufficient conditions of resolution, contrast, etc.

The demand for this intelligent scan is evoked also by the processing method. The straight-line scheme of processing —— input → preprocessing → feature extraction → classification —— may have worked well in character recognition, but this conventional method consisting of a sequence of transformations is inadequate for processing complex pictures. In contrast, in the advanced picture processing as shown in Fig. 2-1, while raw data are reduced to higher-level descriptions, the feedback process is important as well. Results in each intermediate stage are fed back to the previous stages, to check if they are correct, to modify the process, or to extract more precise information based on the results so far obtained. A broad area may be first scanned with low resolution to obtain rough information, and then, zooming up to the relevant portion of the picture, the scan may be confined to a smaller region with high resolution.

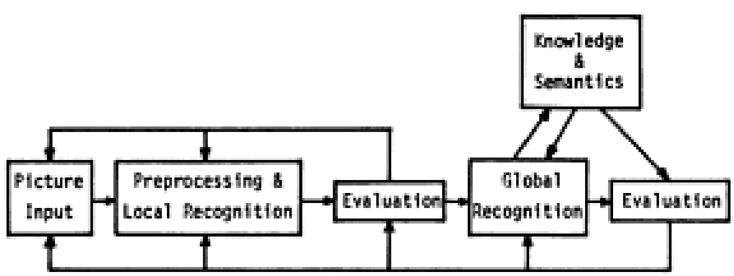


Figure 2-1 Advanced picture processing scheme.

For this purpose, a flying-spot scanner with random scanning ability or a TV camera with controllable focus and contrast is a suitable device. In order to make an effective use of the device, high-speed, flexible program control is essential. In addition, it is desirable that simple preprocessings (e.g., noise filtering) are executed in the earlier stages to reduce storage requirement or processing time in the later stages.

It should be emphasized that a picture processing system should be a total system that includes input/output,operations on image data, and higher-order decision making. As shown in Fig. 2-1, an appropriate data structure is to be used to represent "knowledge and semantics", and problem-solving techniques are necessary to control the whole process in a goal-directed manner. These points suggest that file, symbol-manipulation language, e.g., LISP, and methods of artificial intelligence are to be available in combination with usual processing techniques of imagery.

II-2-2. System-Subsystem Organization

Now the requirements have been clarified. To fulfill them I have adopted system-subsystem organization in the design of GIRLS: a high-speed minicomputer controls all the picture input/output devices; they all form a subsystem or satellite of a host machine. This type of organization is not new. It has been used in several computer graphics systems[9] and picture processing laboratories[15][25].

By organizing the system in this manner, it has several advantages:

(1) The picture I/O devices are connected with the minicomputer via rather simple interfaces, and their control is software-oriented. This results in highly flexible control and parameter setting.

(2) Special resident programs in the minicomputer allow high-speed control of devices and immediate response in man-machine interaction.
(3) Modification of device control is accomplished merely by changing the program.
(4) Among the many picture processing capabilities provided by the system, some of simple preprocessings and feature measurements are performed in the satellite, as well as intimate man-machine interaction. This data reduction in the satellite results in efficient use of the data channel and analysis power of the central computer.
(5) The satellite can share units of disk, magnetic tape and line printer with the central computer. This avoids double investments on these expensive peripherals.

The components, functions and software of the system will be described in the succeeding sections.

II-3. System Components

Fig. 2-2 shows the hardware structure of the new picture processing system: the central host machine is NEAC 2200/200 with several standard peripherals; the GIRLS subsystem consists of a minicomputer MACC 7/F combined with a flying-spot scanner, a storage-type CRT, a character CRT and a Graf/pen (tablet input device). An overall view of GIRLS is shown in Fig. 2-3.

NEAC 2200/200 is a medium-sized character-oriented machine. The current configuration contains 32 k 6-bit-characters of core (2–μsec cycle time), line printer, disk, five magnetic tape units, paper tape

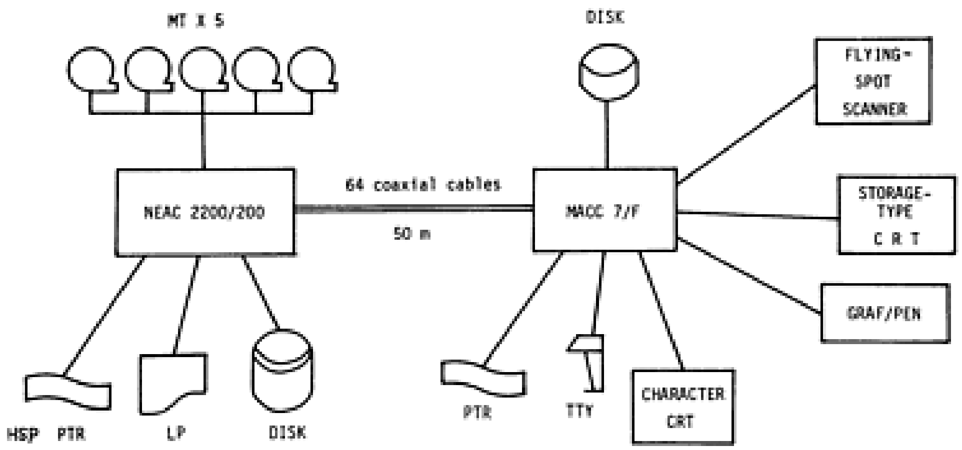


Figure 2-2 Hardware structure of picture processing system by computer complex. Right side is the GIRLS subsystem.

reader and high-speed puncher. In terms of programming, the features of variable-length word, 2-address instruction and implicit addressing allow programmers to write efficient picture processing programs. LISP language is also available. So far an efficient Laplacian-operator program, a region partitioning program, a program for analysis of human-face photographs, a program for matching line figures and others have been implemented in the laboratory.

MACC 7/F is a high-speed minicomputer with 8 k words of wire memory (16 bits/word and 0.6-μsec cycle time) and is equipped with a small disk device and a character CRT. The disk is used to store various program packages, so that they are loaded and run when necessary. The capability of high-speed calculation enables the minicomputer to process image data while controlling the I/O devices simultaneously.

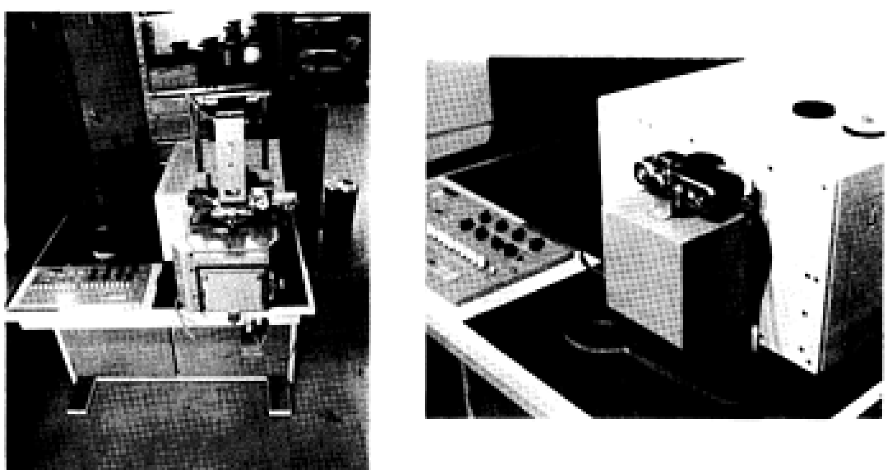
The GIRLS-interface controller consists of specially designed interfaces to combine picture I/O devices with MACC 7/F and to connect

Figure 2-3 View of GIRLS subsystem: (1) I/O typewriter; (2) Paper tape reader; (3) CPU of MACC 7/F mini-computer; (4) Graf/pen; (5) Transmission cable to NEAC 2200/200; (6) Storage-type CRT; (7) GIRLS-interface controller; (8) Flying-spot scanner; (9) Character CRT; (10) Mini-disk device

MACC 7/F with NEAC 2200/200. The controller has special features in order to allow MACC 7/F to control flexibly the attached devices and to act as an intelligent terminal of NEAC 2200/200. (See APPENDIX A-1)

The flying-spot scanner(FSS) is OS-70J of Kowa Co. Ltd. with some home-made modifications(Fig. 2-4). It scans both opaque (75 mm x 75 mm) and transparent (10 mm x 10 mm ∿ 24 mm x 24 mm) materials in a random scan mode; X-Y location of the spot is specified by 10 bit coordinates by the program, in total 1024 x 1024 addressable points. The 5" CRT (C-6861P46; Fujitsu Ltd.) uses a new phosphor material, yttrium silicate(Y_2SiO_5:Ce). Its output spectrum is similar to that of P16, but it has much better properties than P16: shorter rise time, about ten times as bright as P16, and much longer lifetime. The spot size on the tube (70 mm x 70 mm raster) is approximately 35 μm, thus making all the addressable points resolvable. The output signal is digitized into a 5-bit-precision number (i.e., 32 gray levels). Contrast and offset of the gray level are changeable in 4 different modes by a program-settable parameter. The scanner is equipped with a film transport which advances or reverses the film by a pulse motor. The FSS is also able to be used as a film recorder to produce a half-tone picture by readily replacing the scanner head with a camera head as shown in Fig. 2-4(b).

Graf/pen is a tablet with a stylus pen for coordinate input. Either a 14" x 14" or 20" x 20" tablet can be used. The stylus combines a ball-point pen with a tiny spark gap providing a very-low-energy spark which generates a sound pulse. The position of the stylus is detected by measuring the time for the sound to propagate



(a) (b)

Figure 2-4 Flying-spot scanner : with (a) scanner head; or (b) camera head attached.

from the pen to the X and Y sensors along the X and Y axes of the tablet. Data acquisition rate is up to 200 coordinate pairs per second with the resolution of 1024 x 1024 line pairs.

Storage CRT display unit (Tektronix TYPE 611) is a data storage and display instrument. The dimension of the display frame is 16 cm x 16 cm, and 1024 x 1024 points are addressable. Because refreshing is unnecessary, a storage CRT provides an economical means for presenting information in a visual form at the terminal.

II-4. Functional Description of the System

Fig. 2-5 depicts a basic concept of the developed system from a functional point of view. The satellite, with MACC 7/F in its center, offers a set of rather simple but versatile functions: flexible scanning of pictorial data by means of an FSS, preprocessing (e.g., enhancement, noise cleaning, differentiation, etc.), feature measurements (e.g., area, moment of the component, histogram of gray levels, etc.), display of intermediate or final results on CRT, man-machine interaction via TTY or character CRT, and so on. In the host NEAC 2200/200, on the other hand, the researcher can write more problem-oriented program which handles the major data structure and controls the main sequence of picture analysis. The functions provided by the satellite are combined for use when necessary; the program in NEAC can call a MACC routine by such a macro-command as "Scan the film in such and such manner, preprocess the data, and send back the results." These capabilities are sure to shorten the developmental time of a new picture processing program.

Another important feature of this system is that the satellite is able to perform an intelligent job for itself, accessing to the

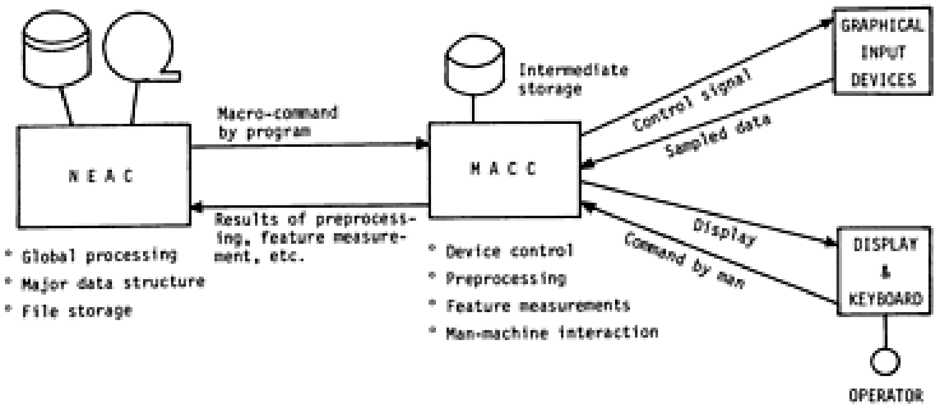


Figure 2-5 Concept of the system organization from a functional point of view.

host machine from time to time when it needs file or larger, complex support. A typical example would be processing of a long film of time-varying patterns, e.g., the motion film of amoebae. The satellite continues to observe the movement of amoebae by processing the film successively, and only when significant or interesting change appears, it interrupts the host, sends it the data and requests further processing which may require a larger data structure.

To support computer-computer communication between NEAC 2200/200 and MACC 7/F, two special programs have been developed; α-monitor in MACC and β-interface routine in NEAC. They handle the demands for a routine set-up from NEAC to MACC and the demands for a file access from MACC to NEAC. Under the control of α-monitor and β-interface routine, the data transfer is extremely simplified by macro-instructions that enable MACC users to use the units of disk and magnetic tape of NEAC in exactly the same way as though they were attached to MACC.

The minicomputer MACC 7/F functions as a switcher of data flow among the I/O devices and NEAC. Sampled data from FSS may flow to NEAC or CRT. Computer output from NEAC may be displayed on CRT, typed out on TTY, or recorded on film by means of FSS. The arbitrary data flow is established just by specifying the destination device, by a program command or an operator's command, before transfer of one block of data. While going through MACC, data is often subject to certain processing and transformation. In this sense, the minicomputer plays three functions: switching, common memory of I/O devices and processing.

II-5. Various Facilities Provided by the System

This section describes facilities which the developed system provides. APPENDIX A will give the detailed technical description of hardware and software which have been developed to realize those facilities.

II-5-1. Flexible Scanning Program of Flying-Spot Scanner

Among the versatile capabilities offered by GIRLS, one of the most important is the special provision to permit flexible scanning of pictorial data by means of a flying-spot scanner.

The FSS can scan and digitize the brightness information of prints or films in a programmed manner within 15 μsec/point: see APPENDIX A-1-2. The scanning program combines this data sampling with data processing in a very flexible manner. The basic idea is to realize any combination of beam intensity (contrast and range of gray levels), spot (shape and operation of scanning light beam), raster (place and way of moving the spot in the whole 1024 x 1024 raster), transformation of data (e.g. binary) and manipulation (feature measurement, storage, or display). Fig. 2-6 shows the repertoire so far available for each category. All these functions except the beam

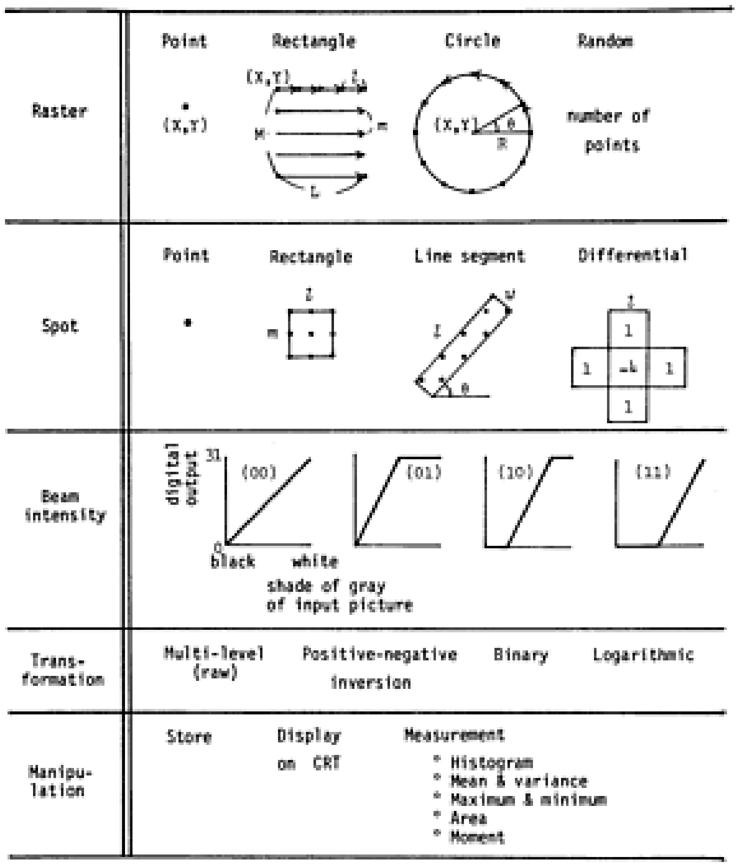


Figure 2-6 Flexible scanning program ; structure and repertoire.

intensity are realized in terms of software, not by specialized hardware, taking advantage both of the random sampling capability of the FSS and the high-speed calculation of MACC 7/F.

Take, for instance, the combination of rectangular raster, the spot of line segment, log transformation and measurement of histogram with all the necessary parameters specified. A sequence of logarithmic gray levels is obtained which a slit-like spot would produce if it moved in the manner of TV scan within the specified area. The assumed output of a slit spot is calculated by sampling the points inside of the slit area and averaging their output value. A histogram of the gray levels is obtained as a final output. A large volume of scanned data may be transferred to the disk memory of NEAC, stored in MT, and/or displayed on CRT, according to the specification.

When the specification of scanning is given by the command sent from NEAC program or entered by an operator at the satellite, a corresponding scanning program is generated; as shown in Fig. 2-7, routines of the specified spot, transformation, and manipulation(s) are inserted in the specified raster routine. Fig. 2-8 shows a typing format for the specification of scanning; underlined characters are the system output. In this way, our FSS is practically given a variety of abilities which originally it did not have, but which are valuable in picture processing, e.g., variable size and shape of spot, linear, binary output, and circular raster. Some of preprocessing operations such as averaging, and differentiation, and feature measurements are performed as the material is being scanned.

Fig. 2-9 through Fig. 2-12 contain some examples of usage of the flexible scanning program, which demonstrate its usefulness and flexibility. The two-bit number of beam intensity selects one of four relationships between digitized output and brightness of the material to be sampled. For the beam (00), 32 quantization levels are spread uniformly over the whole range of brightness, whereas for the beam (01), they are concentrated into the darker range, and the

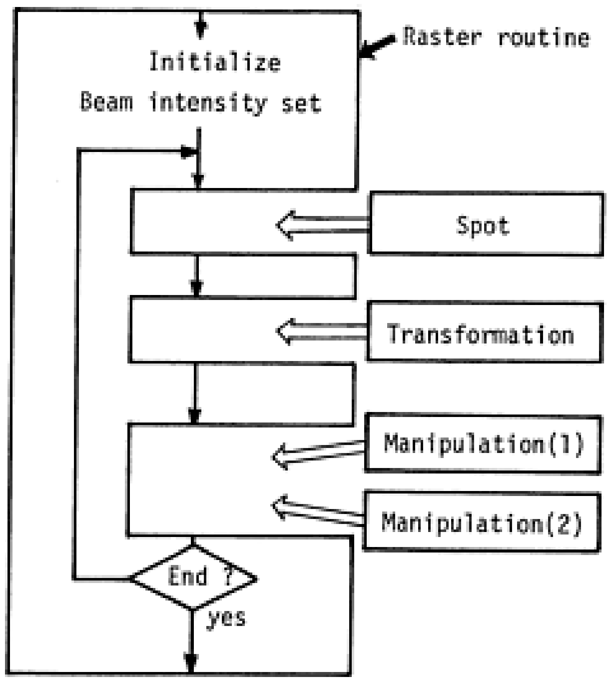


Figure 2-7 Generation of a scanning program.

```
:PROGRAM NO.=0

BEAM INTENSITY=0
RASTER NO.= 1 RSX=310 RSY=120 RAA=120 RBB=208 RXL=3 RYL=3
SPOT NO.= 0
TRANSFORMATION NO.= 0
MANIPULATION NO.= 1
DISPLAY NO.=1
```

Figure 2-8 Typing format of the scanning program.

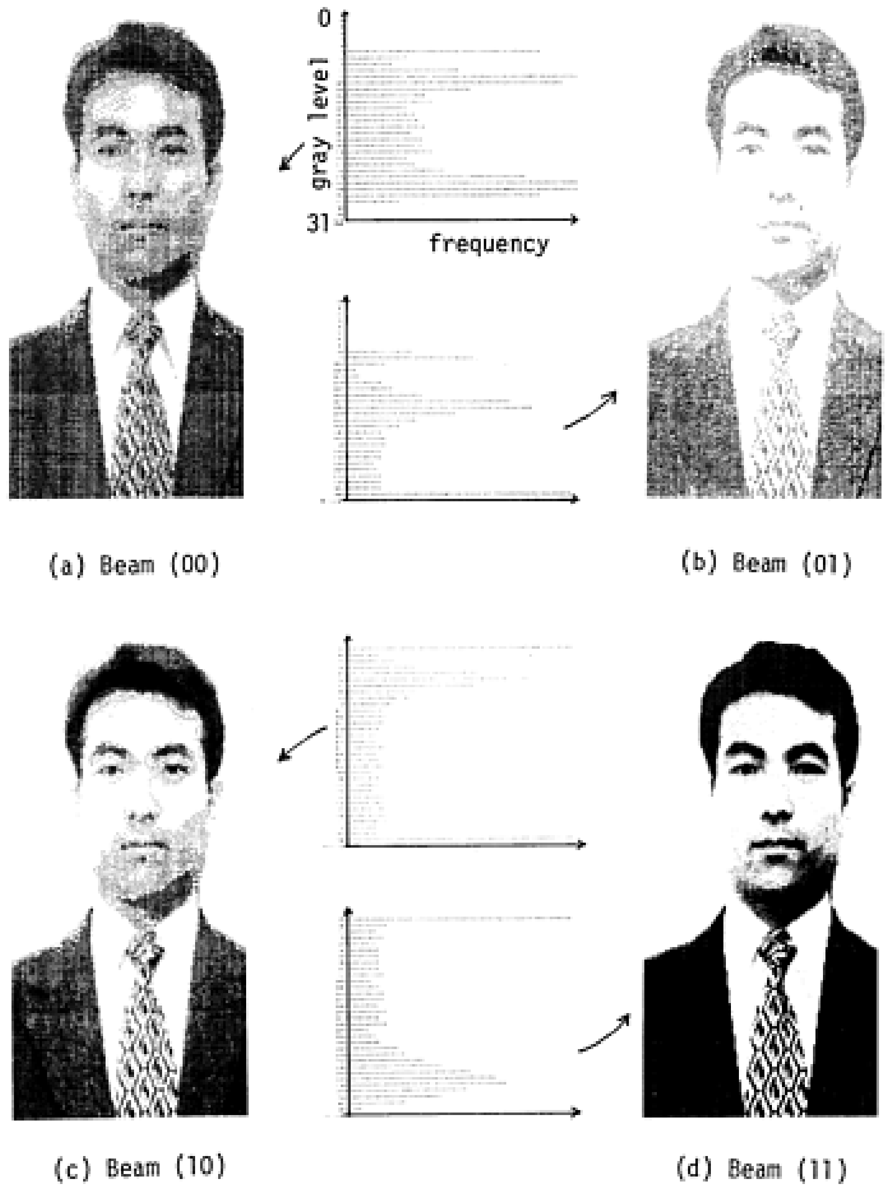


Figure 2-9 Digitization by four kinds of beam intesity. The picture of Fig. 1-1(right) is digitized into 120 x 200 elements. Point spot(1 x 1) is used and the sampling interval is 3. The histograms of gray levels show the effect of the beam-intensity control.

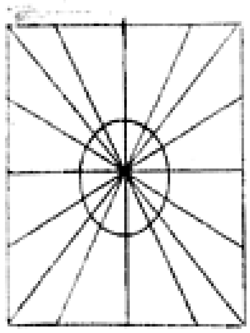
Figure 2-10

Scan by a large spot. The same picture as in Fig. 2-9 is scanned by a 7 x 7 spot. It shows the averaging effect.

(a)

(b)

Figure 2-11 Differential spot
(a) Original picture,
(b) Scan by the differential spot. Edges in the picture are detected.



(a)

(b)

Figure 2-12 Line segment spot
(a) Scan of a line figure by the point spot,
(b) Scan by the line-segment spot. Lines with a particular direction are emphasized.

brighter range gives saturated output 31. The beams (10) and (11) correspond to the middle and brighter ranges, respectively. The brightness digitization, therefore, has substantially 6-bit resolution for gray scale. Pictures and histograms of Fig. 2-9 illustrate the effect of this beam-intensity control.

Fig. 2-10 and Fig. 2-11 present examples of spots which involve preprocessing: averaging by a large 7 x 7 rectangle spot and differentiation by a differential spot to detect edges in a picture. Line-segment spots are very useful in detecting lines with a particular direction as shown in Fig. 2-12. Scan by a 1 x 100 line-segment spot with the slope of 1/2 produced Fig. 2-12(b). Simple thresholding applied to this picture will single out lines and arcs of that particular slope.

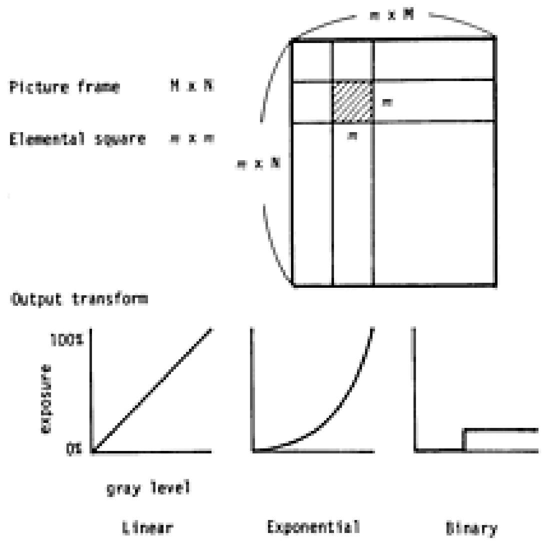


Figure 2-13 Structure of film-recording program.

II-5-2. Film-Recording Program

The flying-spot scanner can be used to record computer output on film, when a camera head is attached in place of the scanner head (Fig. 2-4(b)). The film-recording program has been developed to produce a half-tone picture.

The program follows the basic principle of representing shade of gray by the number of dots plotted in a small square. As shown in Fig. 2-13, the structure and usage of the program are similar to the input scanning program. One can specify the size of picture frame M x N, the size of square of one picture element $m \times m$, and one of the output transforms. The program reads the image data to be recorded on the film out of the disk memory of NEAC.

Experimentally it was proved in our FSS that when the "beam-on" time (exposure time) for one plot is set to 0.6 μsec, superimposed plotting is distinguishable up to eight times with linearly increasing film density produced in the film. Therefore 8 x m x m exposure levels are available for each m x m square which corresponds to one picture element.

The larger the elemental square is, the higher the gray-level resolution is for each picture element, but the lower the positional resolution is.

The output transform is the manner of allotting those exposure levels to 32 gray levels: linear, exponential or binary. The exponential allotment is used to compensate for the logarithmic characteristic of the visual organ. The allotment is given by a table indicating percentage of available exposure levels versus gray level, and therefore an arbitrary one can be realized merely by replacing the table in order to make compensation for any kind of distortion involved in the film manipulation process or perceptual process. Table 2-1 shows the real numbers of plots in the case of 5 x 5 square with exponential transform.

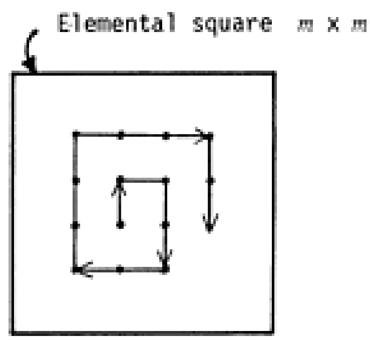


Figure 2-14

Plotting order.

Plotting in the elemental square is performed in the spiral order as illustrated in Fig. 2-14 from the center to the marginal points. Each point is exposed for 0.6 μsec. When all the points have been plotted once, the second superimposition plot is performed, then the third time and so on, until the square receives, in total, the specified number of plots (exposure levels).

gray level	number of plots	gray level	number of plots
1	0	17	23
3	1	19	31
5	2	21	43
7	3	23	59
9	5	25	80
11	8	27	109
13	11	29	150
15	17	31	200

Table 2-1 Plot distribution in the case of a 5 x 5 elemental square with exponential transform.

Fig. 2-15 shows an example of usage of the program. An original picture(Fig. 2-15(a)) is first scanned and digitized into 140 x 200 picture elements by means of the input scanning program. Then a half-tone picture is reproduced using the film-recording program.

A 5 x 5 elemental square is used with linear transform (Fig. 2-15(b))or exponential transform (Fig. 2-15(c)). The size of the elemental square is 350 μm x 350 μm on the tube of FSS. Since all the addressable points on the tube are resolvable, soft focus of the camera is preferable to make the resultant pictures more natural.

Figure 2-15

Example of film recording. The picture frame is 140 x 200 and the elemental square is 5 x 5.

(a) Original print

(b) Reproduced print; linear transform (iris 8.0)

(c) Reproduced print; exponential transform (iris 8.0)

II-5-3. CRT Display of Gray-Level Picture

Our monitor display (storage-type CRT) has a 16 cm x 16 cm square display area consisting of 1024 x 1024 addressable points. Though the CRT can express only black or white at each point, it is very desirable to display half-tone images. For this, the shade of gray of one picture element is represented by the density of white points filled in a 3 x 3 square array on the CRT: we can display 10 gray levels. In this way, a gray-level picture of up to 341 x 341 picture elements can be displayed, with satisfactory resolution and shades of gray for the purpose of monitoring the picture under processing.

II-5-4. Facilities for Intercomputer Communication

α-monitor in MACC 7/F and β-interface routine in NEAC support the intercomputer communication. Their main functions are twofold. One function is to set up a program to run in MACC 7/F. Under the control of α-monitor, all the programs in MACC 7/F are identified by the registered program number. There are three ways for setting up a program to run in MACC 7/F in order to allow free combination of programs which are prepared and registered in MACC: (i) A program in NEAC is able to call a program registered in MACC by sending a calling segment which contains the program number and arguments of the program. This is one of the means for the programs in the two computers to communicate; (ii) An operator at MACC 7/F can input the program number and arguments via TTY or PTR, in order to call the program into action; (iii) A program in MACC can call other routines by executing a special macro-instruction.

The second function is to provide MACC users with facilities for data transfer from/to the core memory, disk, and MT of NEAC. Macro-instructions have been prepared. They enable MACC to share storage devices and a lineprinter with NEAC, and offer ease of programming and economical advantage. (See APPENDIX A-2)

In summary, communication between programs in two computers is accomplished by two means; to send a calling segment and to share a large volume of data in storage devices of NEAC.

II-6. On-Line Analysis of Human-Face Photographs

This section describes the analysis of human-face photographs in on-line mode, as an illustration of a potential use of the developed system. The analysis program was first coded in NEAC assembler language to process photographic data which were stored on a magnetic tape. This older version was converted into the on-line version described below. It took only several days to complete the conversion by utilizing the facilities provided by the above mentioned system.

Fig. 2-16 illustrates the sequence of events that occur in the on-line analysis. An operator at the satellite sets a photograph of a face in the scanner head of the flying-spot scanner. He presses the interrupt button and calls the scanning program into action via α-monitor. The program scans the photograph in the specified manner. The digitized data with an average operation performed are transferred and stored in the disk memory of NEAC. In NEAC, a line-extraction program performs the operation of two-dimensional secondary differentiation, and produces a binary picture which represents the contour of the face by thresholding the differential picture. The binary picture is also stored in the disk. Then, the analysis program for human face begins to process the binary picture and locates the feature points of the face, such as eyes, nose, mouth, chin and so on. CHAPTER III provides a detailed description of the analysis method.

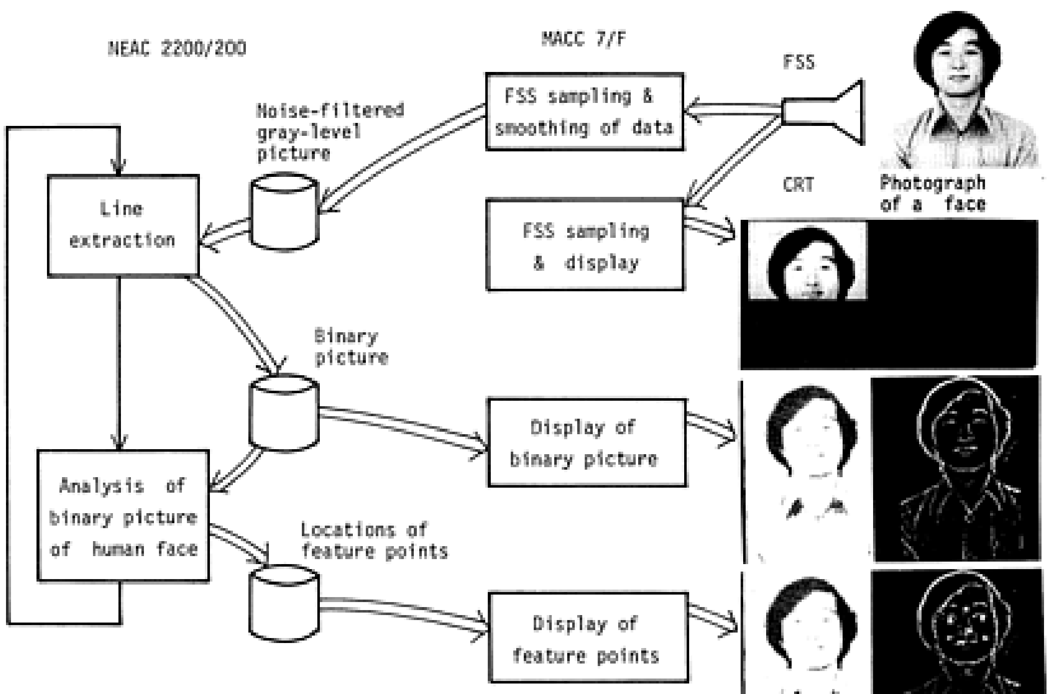


Figure 2-16 Sequence of events in on-line analysis of human-face photograph.

While NEAC is doing these tasks, MACC displays the face which is now being processed on the CRT: on the left half, the gray-level picture is displayed, and on the right, the binary picture whose data are read out of the disk memory of NEAC. When all the feature points have been extracted by the NEAC program, their coordinates are transferred to MACC and displayed as visible points in the binary picture on CRT. This completes one cycle of the analysis. Each routine in MACC is set up to run just when the data it needs have been prepared. The processing time including input and display is 1 $\sim$ 1.5 minutes per a photograph.

II-7. Supplementary Remarks

In this chapter, only a general descriptive presentation of the system was given. For the detailed technical description, see APPENDIX A.

Other than computer analysis of human-face photographs, the system has already been applied to diverse tasks: cloud measurement in the picture obtained from ESSA; pictorial analysis of traffic flow; design and generation of more than a thousand digital KANJI patterns.

This picture processing system will be incorporated in the Inhouse Computer Network[30] which is now under development at Prof. Sakai's Laboratory, Department of Information Science, Kyoto University. When it is completed, the system will be able to enjoy free and on-line use of distinctive facilities such as Fast Fourier Transform, micro-programmed computation ,etc. which other computers in the network provide. The network will greatly enlarge the capability and applicability of the picture processing system.

CHAPTER III

ANALYSIS OF HUMAN-FACE PHOTOGRAPHS

III-1. 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 chapter, the problem of picture analysis of human faces is presented as an example of automatic analysis of complex pictures. A computer program for it has been developed: given a photograph of the human face, the program extracts feature points such as the eyes, mouth, nose, and chin. A flexible picture analysis scheme with feedback is successfully employed in the program.

First, the significance of the problem is briefly described. Then the computer analysis procedure for human faces is outlined and the flexible picture analysis scheme with feedback is introduced. A complete description of the analysis will be given in III-4. More than 800 photographs have been processed by the program and the results are presented. This chapter principally treats picture analysis. The problem of face-identification using the located feature points will be discussed in CHAPTER IV.

III-2. The Problem —— Human Face as Object of Picture Analysis

Human faces are very interesting as an object of picture analysis for several reasons:

(1) They are not artificial, and not as simple as cubes or pyramids which have been used in the visual scene analysis of hand-eye projects.

(2) A face has many component substructures: eyes, nose, mouth, chin and so on, which can be recognized as such only in the proper "context of the face".

(3) These components are distributed in the face within a certain permissible range, whose mutural relations can be correctly grasped by the concept of picture structure.

(4) Lines in a face are very difficult to define, difficult to extract and are not straight, all of which make the problem of interest.

(5) The variety of human faces is as large as the human family.

Pictures of human faces contain many essential problems to be investigated in the field of picture processing. How far the automatic analysis procedure can go into detail and can categorize the human faces has interested me.

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. The problem in this chapter is, given a photograph of the human face, to locate the face-feature points in order to characterize the face.This kind of problem is referred to as feature extraction in pattern recognition, and is regarded as most difficult and important. In the attempts at computer recognition of faces made by Bledsoe[6], Kaya and Kobayashi[16], and Goldstein, Harmon and Lesk[12], the feature extraction was due to human-operator's ability to extract or describe features of the face. In contrast, the works by Sakai, Nagao and Fujibayashi[31], Kelly[17] and Fischler and Elschlager[8] are attempts to automate this feature extraction process, as is also the case with my work reported here.

III-3. Outline and Features of Analysis Method

This section outlines the analysis method of human-face photographs to give a general view of the program. Then is explained the picture analysis scheme with feedback, which is employed in the program.

III-3-1. Outline of Analysis

Fig. 3-1 is the block diagram illustrating the flow of analysis. Main body of the analysis program was written in the NEAC 2200/200 assembler language.

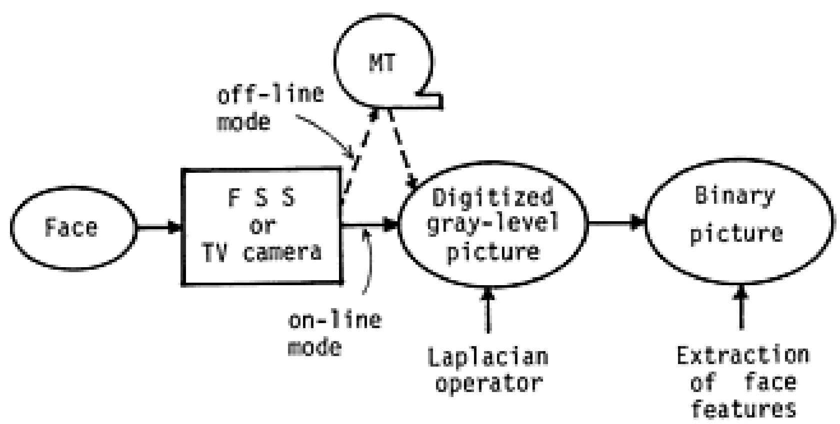


Figure 3-1 Block diagram of analysis of human-face photographs.

(a) Original photograph (b) Printout of the digital gray-level picture (c) Binary picture

Figure 3-2 Picture input and line extraction. The dark horizontal line in the upper part is due to the burn in the CRT surface of the FSS used for digitization.

A photograph (or a real view) of the face is converted to a digital array of 140 x 208 picture elements, each having a gray level of 5 bits (i.e., 32 levels), by means of a flying spot scanner or a TV camera. Fig. 3-2(a) is an original photograph, and Fig. 3-2(b) is the gray-level picture digitized by the FSS. The program can work in either on-line or off-line mode. In on-line mode, the digitized picture is fed directly into the analysis program, and in off-line mode, it is stored once on a magnetic tape for later use.

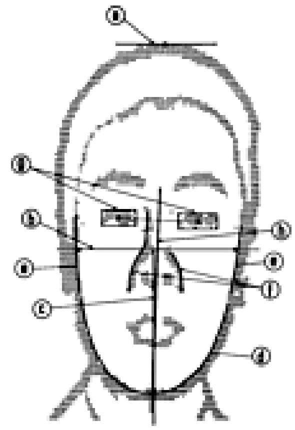
The binary picture of Fig. 3-2(c) is obtained by applying a Laplacian operator (i.e. two-dimensional secondary differentiation) on the gray-level picture and thresholding the result at a proper level. This binary picture represents contour portions where brightness changes significantly. The analysis program which locates face-feature points works on this binary picture.

The analysis steps are first to find the approximate position of the face, and then to go into detail on the positions of face

Figure 3-3

Typical sequence of the analysis steps.

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



components. These steps are shown in Fig. 3-3 from(a)to(h). There are several backup procedures with feedback among them, hence some analysis steps might be tried more than once. Various picture processing techniques are used in the analysis program.

Figure 3-4

Result of feature extraction.

As a final result, more than thirty feature points are located in the binary picture as shown in Fig. 3-4.

Once these feature points are located, much more precise features can be measured by returning again to the original photograph. The method and significance of this refinement process will be discussed in CHAPTER IV.

III-3-2. Flexible Picture Analysis Scheme with Feedback

The distinguishing feature of the 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 backup procedures with feedback.

As shown in Fig. 3-5, each step of analysis consists of three parts: prediction, detection, and evaluation. The result is evaluated for the purpose of knowing whether the program can proceed to the next step. If an unsatisfactory result is yielded, its cause is examined to find out the step where the analysis must be performed again. There are two cases of feedback. Direct feedback is used to modify the parameters of the analysis step just executed. The other is a feedback to former steps. It takes place when the imperfection of the result is disclosed during the succeeding analysis in the form of an unsatisfactory evaluation; usually the evaluation criteria are not so severe, and so sometimes the process may have gone to the next step even though the analysis at one step is not perfect. In this way, the total process of analysis becomes flexible and adaptive. The analysis in a certain portion of a picture may be performed several

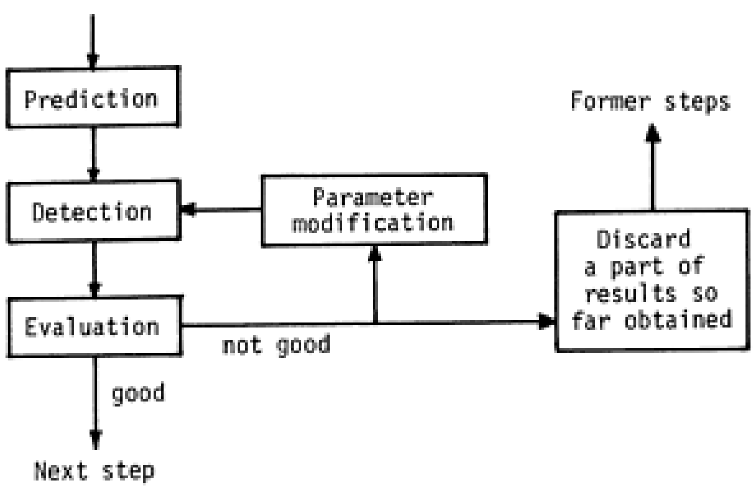


Figure 3-5 Structure of analysis step.

times with a change of parameters until it gives a satisfactory result.

I do not claim that this scheme is completely new. Similar ideas may have been suggested by several authors. I believe, however, that this is one of the first works that actually implement the scheme with feedback and show its powerfulness in the picture analysis.

This analysis scheme has the following excellent properties:

(1) Inevitably the analysis proceeds in a top-down goal-directed manner, which makes the analysis efficient because the necessary local operations are applied on only the relevant portions estimated by the process.

(2) The analysis algorithm which is to be prepared for each step can be fairly simple, because it knows what to look for and can be repeatedly performed on the most probable area with improved parameter values using the feedback process. Unlike the straight-line scheme, errors at each step are not accumulated, but will be disclosed in the succeeding steps and recovered by the backup procedures. This is one of the most important features.

(3) The economy of memory and processing time for the analysis is great if a flexible scanning device is available which samples only the significant parts. This advantage is fully taken in the extraction of precise locations of feature points, which will be described in CHAPTER IV.

III-4. Complete Description of Analysis

This section describes in detail the flow of analysis that is illustrated in Fig. 3-1; from input of picture through final result.

III-4-1. Input of Picture

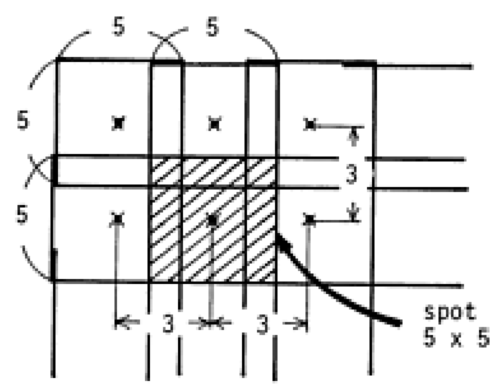
A picture (or a real view) of a face is first transformed into a digital array by means of a flying-spot scanner(FSS) or a TV camera. When the FSS is used, the standard material is a photograph in which lies a full face of about 2 cm. The photograph is scanned by the flexible scanning program described in II-5-1, and digitized to an array of 140 x 208 elements, each of 5 bits of brightness information (i.e. 32 levels). As shown in Fig. 3-6, a 5 x 5 rectangle spot is used and sampling is made at every 3 points. The face in the digitized picture is typically of size 80 x 80 ∿ 100 x 100. The necessary time for input is about 5 seconds. Fig. 3-7 shows a line printer output of a digitized picture with various characters superimposed to approximate the gray levels.

The reason why the input digitization is rough sampling by a rather large-sized spot is that extreme details are not necessary in the next line-extraction process: averaging effect of a large spot suppresses high-frequency noise and thus prevents extra unimportant lines in a face from being extracted.

Since the scanning program of the FSS is highly flexible, appropriate commands of an human operator allow various kinds of photographs to be digitized into the standard format: fine scanning for a smaller-sized photograph and coarse scanning for a larger-sized one; furthermore in the case of a photograph in which there are several people, the scan and input of one particular face can be performed very easily.

Figure 3-6

Scanning by a flying spot scanner. The spot size is 5 x 5, and the sampling interval is 3.



When a TV camera is used as an input device, the human operator adjusts the camera watching a monitor TV, so that the object face may be seen at the appropriate position with the appropriate size. The video level is also adjusted manually. The whole frame consisting of 240 x 300 picture elements is completely digitized and then it is edited into an array of 140 x 208 by software.

The picture digitized in these ways is either fed into the face analysis program in on-line mode, or stored on the magnetic tape in the data collection stage of off-line analysis.

III-4-2. Line Extraction by Laplacian Operator

The next step of processing is to extract lines and contours from the picture. The operator illustrated in Fig. 3-8 is applied to the digital picture of Fig. 3-7, and then thresholding the result at a proper threshold value produces the binary picture shown in Fig. 3-9. This binary picture represents contour portions where brightness changes significantly. It preserves the structural information of the original gray-level picture and yet far easier to deal with because of the binary values. All the succeeding analysis is performed upon this binary picture. The threshold value is fixed to 30.

As to edge or contour detection, a number of papers have appeared: for a survey, see [22] that contains a large number of references. The operator used here is fundamentally a digital Laplacian operator (two-dimensional secondary differentiation). Let I(x,y) be a function that represents a gray-level picture. The Laplacian is defined as:

$$\nabla^2 I(x,y) = \frac{\partial^2 I(x,y)}{\partial x^2} + \frac{\partial^2 I(x,y)}{\partial y^2} \quad . \qquad (3\text{-}1)$$

The simplest digital version of this Laplacian is

$$\nabla^2 I_{ij} = 4I_{ij} - (I_{i-1j} + I_{i+1j} + I_{ij-1} + I_{ij+1}) \; , \qquad (3\text{-}2)$$

Figure 3-7 Digital gray-level picture of a face.

			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			

Figure 3-8

Operator for line extraction. This is fundamentally a Laplacian operator.

where $\{I_{ij}\}$ denotes the digital array to represent $I(x,y)$. From this equation it can be seen that the operator of Fig. 3-8 combines differential operation with averaging: i.e., first averaging operation is performed on 3 x 3 subarea, then differential operation regarding each subarea as one picture element in the equation (3-2). This operator was demonstrated in Sakai, Nagao and Kidode[32] to work very successfully for line extraction of human-face photographs.

Figure 3-9

Binary picture.

a	b	c
d	e	f
g	h	i

(a) $\sqrt{(a-e)^2+(b-d)^2}$

(b) $\max[\ |a-e|,|b-e|,|c-e|,|d-e|,\ |f-e|,|g-e|,|h-e|,|i-e|\ ]$

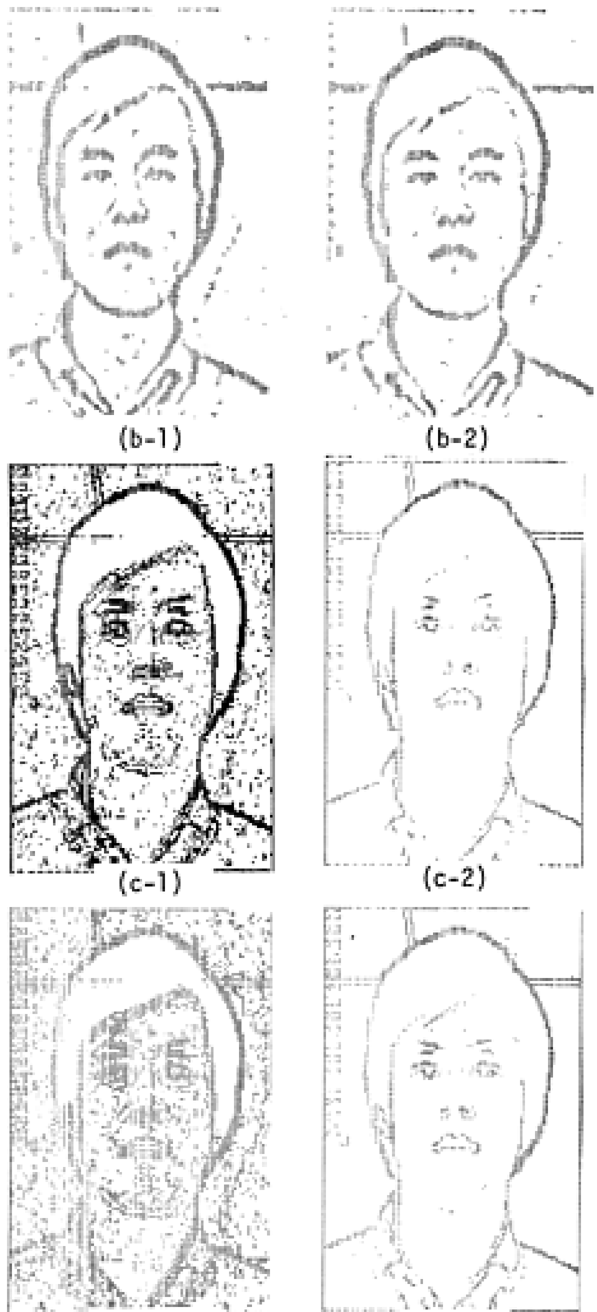
Figure 3-10 (a) Robertz operator; and (b) Maximum of differences.

For comparison, three operators for edge detection including the Laplacian of Fig. 3-8 were applied on the same set of pictures; the other two are the first derivative of Robertz[28] (see Fig. 3-10(a)) and the maximum of differences (see Fig. 3-10(b)) for 3 x 3 window. The results are presented in Fig. 3-11 and APPENDIX B contains supplementary data. They show the superiority of the Laplacian operator I used. It can detect slow but apparent changes in brightness as well as sharp edges, because it takes broader area into account. Furthermore, it is observable that the different values of threshold do not yield any significant difference in the resultant binary pictures.

A very efficient program of the Laplacian operator has been coded, by taking maximum advantage of variable-length word and implicit addressing of NEAC 2200/200, though the computer is not so fast: 2-μsec memory-cycle time for a 6-bit character. It takes only 14 seconds to convert from Fig. 3-7 to Fig. 3-9.

Any additional operations such as thinning, elimination of isolated points, etc. are not performed on the binary picture. The reason is that significance of line elements may vary from position to position, or in other words may depend on the "context": a set of line elements forming a long line may mean the contour of the face or meaningless edges in the background; a group of isolated points may come from noise or show the existence of low-contrast edges around the chin. The "context" information is used positively in the picture analysis program.

(a)



(b-1) (b-2)

(c-1) (c-2)

(d-1) (d-2)

Figure 3-11

Comparison of line-detection operators.

(a) Gray-level picture

(b) Laplacian operator
(b-1) $\theta = 25$
(b-2) $\theta = 35$

(c) Robertz operator
(c-1) $\theta = 2$
(c-2) $\theta = 4$

(d) Maximum of differences
(d-1) $\theta = 2$
(d-2) $\theta = 4$

III-4-3. Analysis Program

Fig. 3-12 illustrates the general flow of analysis: it shows 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 information to detecting accurate positions of eyes, nose, mouth and so on. Each subroutine shown by the square box tries to detect each component (e.g., eye) satisfying specific conditions (e.g., location, size and shape) in the predicted region. A typical sequence of analysis steps is from (a) to (h) as shown in Fig. 3-3.

There are several backup procedures with feedback interwoven among the analysis steps. According to whether one step succeeds or fails in the feature detection, which subroutine to be executed next is determined, and various parameters in the program are modified. The main part of backup procedures will be described in the next sub-section III-4-4.

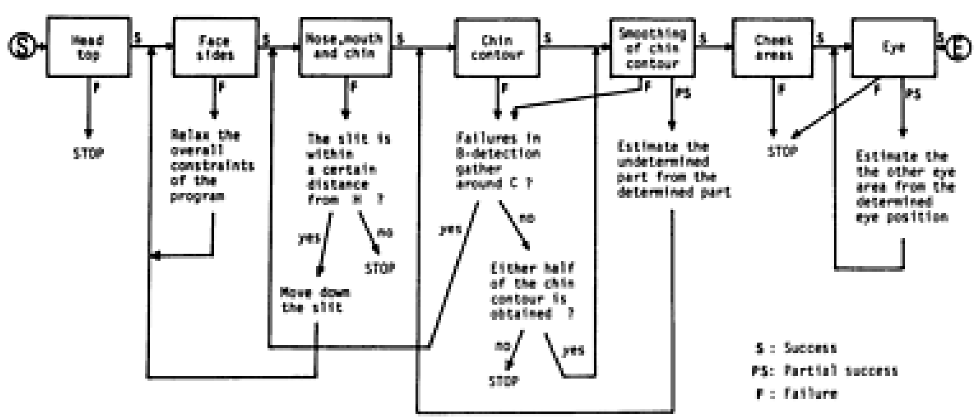
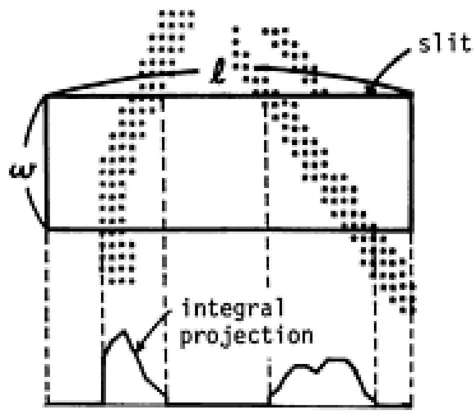


Figure 3-12 General flow of analysis program.

Figure 3-13

Integral projection of a slit.



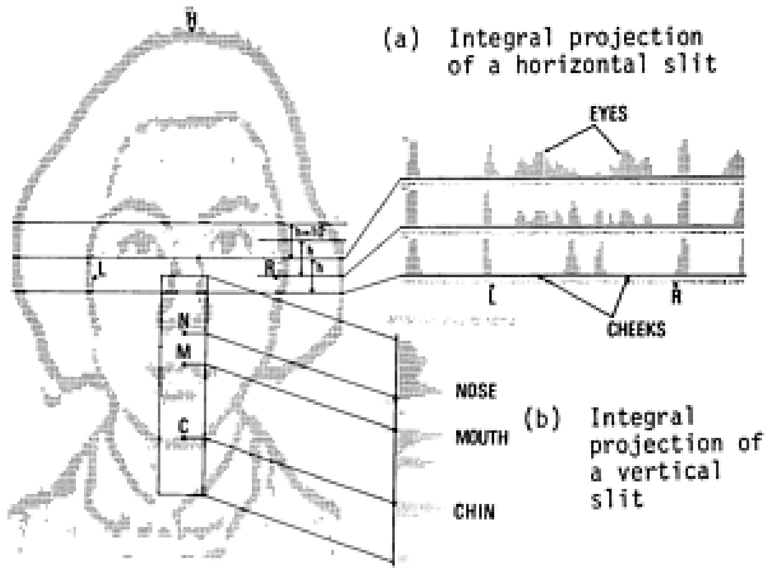


Figure 3-14 Detection of the face sides, nose, mouth, and chin by the application of slits.

A fundamental, useful technique of picture processing used throughout the program is an "integral projection". As shown in Fig. 3-13, 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 is 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 in detail.

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 top of the head 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. 3-14. Fig. 3-14(a) shows three successive outputs. When the slit crosses the cheeks, its integral projection displays characteristic patterns like the lowest one of the three in Fig. 3-14(a): two long clearances and one or two peaks sandwiched by them. The two long clearances are the cheeks, and the outputs at these positions in the upper slits show the eyes or sometimes the effect of glasses. Thus we can determine the left and right sides of the face as the left and right ends of these two vacant portions. They are indicated by L and R in Fig. 3-14. Since we need to know these positions only approximately, the width h of 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. (See example(1) of III-4-4)

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. 3-14. 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. 3-14(b) is one of the typical curves. The integral projection curve is then coded into a compact form by using the symbols for the types of the peak and its length, as shown in Fig. 3-15. The integral projection of Fig. 3-14(b) is coded as follows:

$$R(17)\cdot B(8)\cdot M(5)\cdot B(3)\cdot M(3)\cdot B(10)\cdot M(4) \qquad (3\text{-}3)$$

Before coding, the curve is smoothed so that small gaps or noisy peaks are eliminated.

The form of this curve changes widely depending on the face and existence of face inclination, shadow, mustache, wrinkles and so on. At present, nine typical curves are prepared as standard types. They are also coded in the same manner and stored in the table. The table is shown in Fig. 3-16 with the explanation of how it is read. In a word, each standard type specifies mutual positional relations of the nose, mouth and chin. The table can be extended simply by adding the new types. The integral projection of the input picture is compared with the 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. For instance, the integral projection(3-3) of Fig. 3-14(b) matches TYPE 3. If there are more than one match, the matching criteria of the standard types are changed with a little more rigidity. In case there is no match, the matching criteria are slightly relaxed, or the vertical slit is shifted a

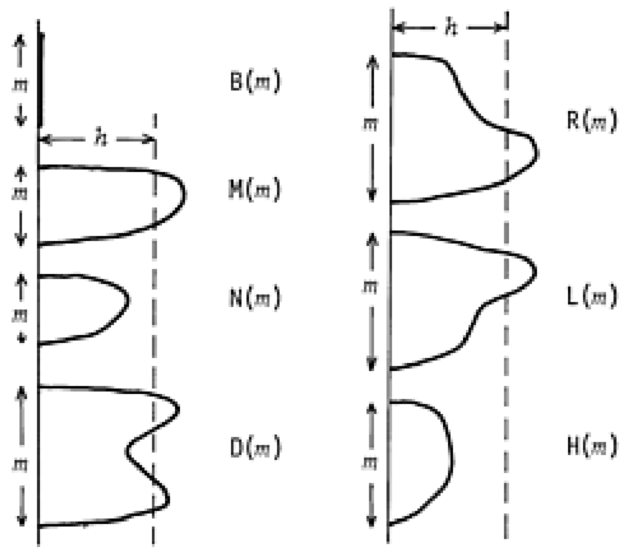
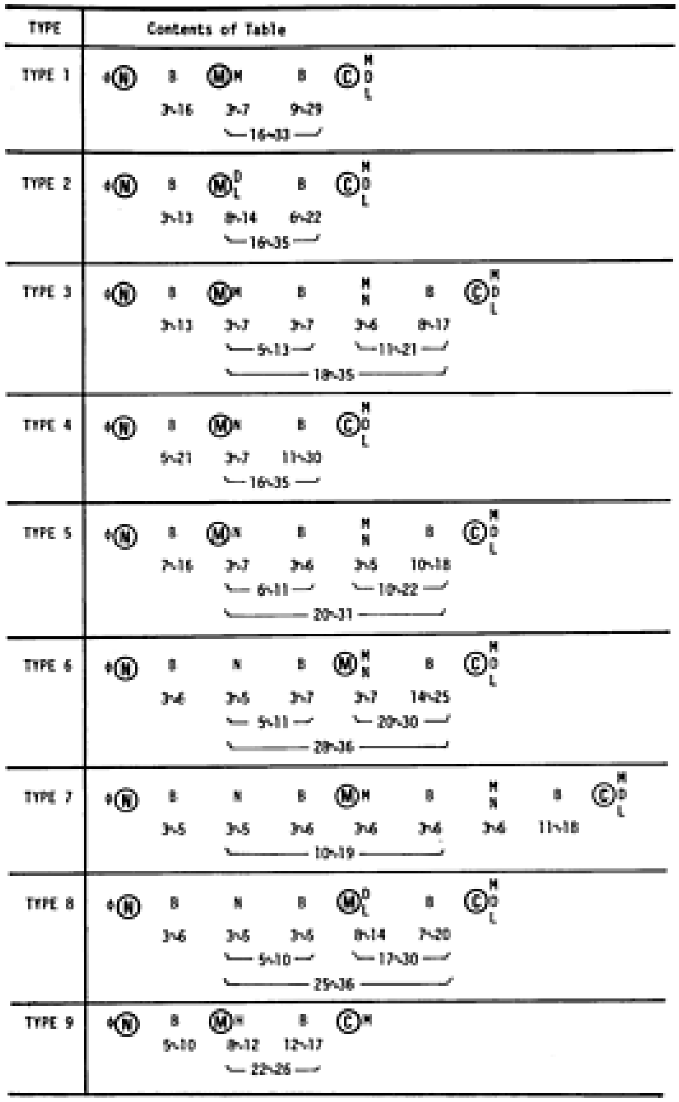


Figure 3-15 Types of peaks.

little. And then matching is tried again. If there is no match even then, this step is a failure, and the backup procedure will work to reactivate the step 2).

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 an elaborate means such as line extension is 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 area for the chin contour. As shown in Fig. 3-17, the search area is determined by L, R, C, M and N. Nineteen radial lines are drawn downward from M in every 10 degrees. These lines are expected to cross the chin contour in the predicted area at nearly right angles, if the



The first line indicates the order of peaks to be satisfied. The second line specifies the length of each peak which appears in the first line. The third and fourth lines indicate the restrictions on the sum of the lengths of successive peaks. Ⓝ, Ⓜ and Ⓒ stand for the locations of the nose, mouth and chin, respectively. φ means " don't care ".

Figure 3-16 Standard types of integral projections.

positions of L, R, C, M and N are correct. A slit is placed along each radial 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 radial lines is smoothed. The 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 radial line whose incidental integral projection contains no conspicuous peak within the search area. In case B's cannot be determined on three consecutive radial lines, the contour detection is judged unsuccessful. The program activates the backup procedure. The cause is investigated and an appropriate recovery action will be taken. (See examples (2) and (3) of III-4-4)

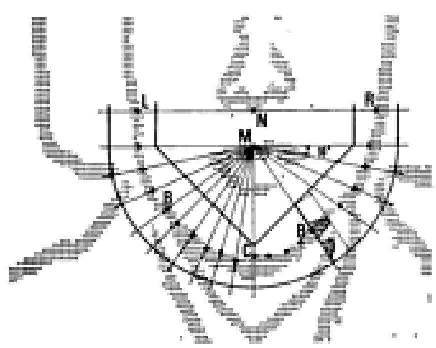


Figure 3-17

Extraction of chin contour. The search area is established and a slit is placed along each radial line.

5) Nose End Points and Cheek Areas

As is shown in Fig. 3-18, by successive application of horizontal slits starting from N, we can trace upward the nose as well as the left and right face sides up to approximately the eye position. Both ends of the nose, P and Q, are located. Then the left and right cheek areas are determined as those which are enclosed by the nose lines and face-side lines.

6) Eye Positions

Examination of the vertical and horizontal integral projections of the cheek areas gives the rectangles containing the eyes. Fig. 3-18 illustrates the method: first, vertical summation is done between the face side and the nose side, and then horizontal summation in the zone obtained from the vertical summation. As in the case of Fig. 3-19(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[29] are executed in the rectangle cut out of the picture to eliminate gaps and hollows which often appear in the eye area(Fig. 3-19(b), (c)). The connected component of the maximum area is then identified and singled out in the rectangle(Fig. 3-19(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. 3-19(e)). The position, the size, and the shape of the component are examined whether they satisfy certain conditions which the locations and properties of feature points so far obtained impose.

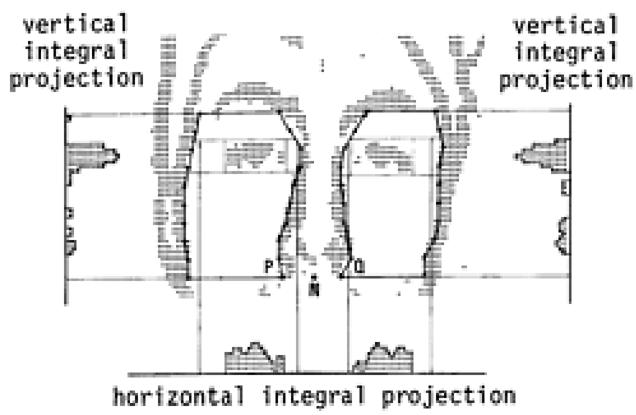
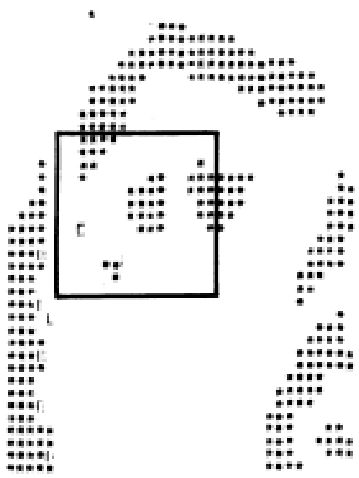
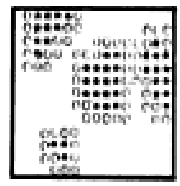


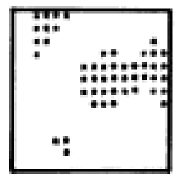
Figure 3-18 Cheek areas and rectangles which contain the eye.



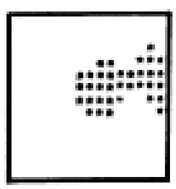
(a) Predicted rectangle



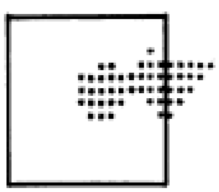
(b) Fusion



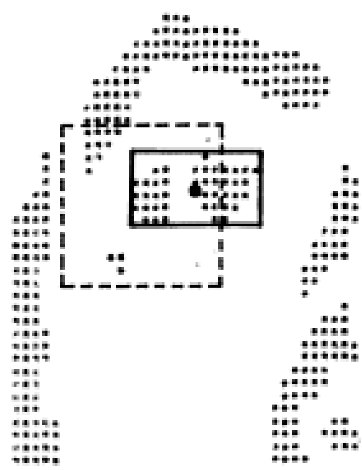
(c) Shrinking



(d) Component of the maximum area



(e) Extended component



(f) Rectangle that circumscribes the component, and its center

Figure 3-19 Determination of eye center.

If they do, the eye center is determined as the center of the rectangle which circumscribes the component(Fig. 3-19(f)). In this way, the left and right eye centers, S and T, are obtained.

This concludes a whole analysis procedure. As the final result all the feature points extracted are printed out on the lineprinter as was shown in Fig. 3-4.

III-4-4. Backup Procedures with Feedback

In the process of face analysis described above, errors may occur in various steps. To manage them, several backup procedures with feedback were introduced. In the following descriptions are demonstrated a few typical examples.

(1) When face sides, L and R, are incorrectly located.

L and R are usually determined at the step 2) by detecting the existence of two wide areas of the cheeks on both sides of the nose. The method is very simple, and therefore sometimes they are misrecognized,especially in a woman's face like Fig. 3-20. When the analysis proceeds to the next step 3), where a vertical slit is applied to find the nose, mouth and chin, the integral projection curve differs markedly from what is expected. Thus the analysis fails there. The program goes back to the step 2) once more, and tries to find out another possibility for L and R. Fig. 3-20 is an example in which the correct results were obtained in the second trial. We could determine these positions correctly in this way for many pictures for which the analysis was first unsuccessful.

(2) When the nose(N), mouth(M) and chin(C) are incorrectly located.

There are cases where N, M and C are determined in the step 3) as upward shifted positions as in Fig. 3-21. This trouble is discovered as failures in detecting the chin contour at the step 4). Usually

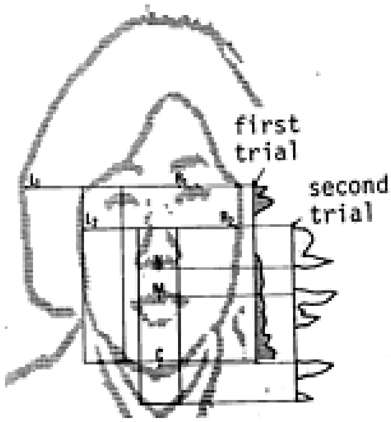


Figure 3-20

Misrecognition of L and R.

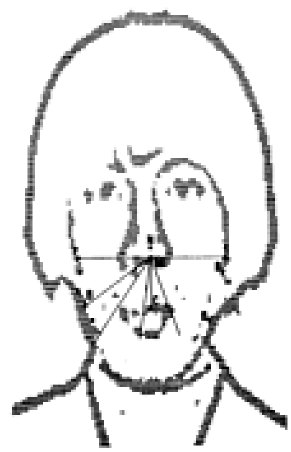


Figure 3-21

Misrecognition of N, M, and C.

failures in B detection occur around the misrecognized C in a symmetrical manner. When this is found, the backup mechanism forces the program to go back to the step 3) to re-examine the positions of N, M and C.

(3) When the chin contour is unsatisfactorily located.

In the detection of chin contour at the step 4), there often exists a short break in the contour, or the real contour line partially sticks out of the search area. Then the complete contour is not determined. Usually in this case, however, either half of the contour is obtained correctly. Thus, as shown in Fig. 3-22, for the radial line whose B was not obtained, a prediction point B' is determined there by utilizing the symmetrical property with respect to the vertical axis. 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.

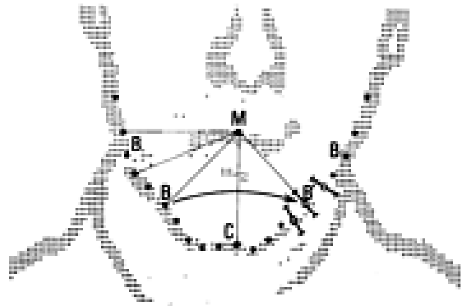


Figure 3-22

Failure in detecting chin-contour points.

(4) When a wrong face-side line results in a wrong prediction area of the eye.

As shown in Fig. 3-23, one of the face-side lines that are determined in the step 5) may extend to the wrong direction, which results in a wrong prediction area of the eye. Thus, at the next step 6) the component expected to be the eye does not satisfy the necessary condition. Then a new prediction area is established from the position of the other eye, and the determination of this eye position is tried again. The face-side line is also corrected by tracing it downward from the correct eye position.

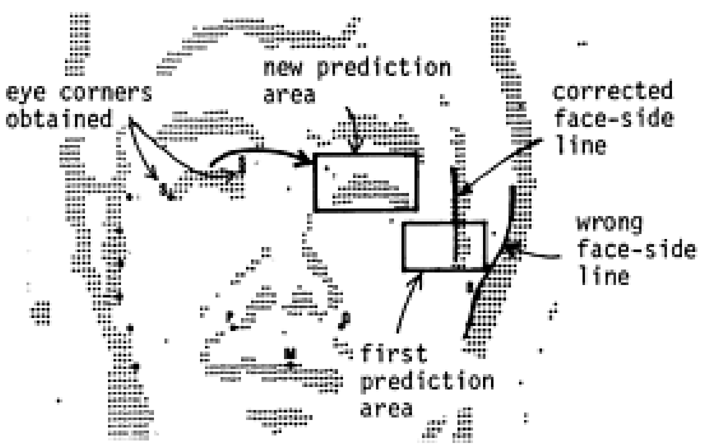


Figure 3-23 Wrong prediction area of the eye.

Other than the examples mentioned above, there are many minor feedbacks. Compound feedback may also occur. For instance, the failure in the chin-contour detection activates the step 4), whose failure, then, makes the program go back to the step 3).

It is notable that this trial-and-error process with the feedback mechanism allowed each analysis step to be simple. The once-and-for-all process would have required that each step be highly elaborate and reliable, in order to avoid delivering erroneous results to the succeeding steps.

In many steps, simple pattern matching methods have been employed. It should be mentioned that pattern matching applied to the restricted region in a picture has a positive meaning in contrast to that applied to 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.

III-5. Results of Analysis

About 800 pictures of human faces have been processed by the program described above. Most pictures (688) of this large data set were obtained in digital form at the World Fair '70 OSAKA in 1970. Nippon Electric Company ran an attraction named "Computer Physiognomy": a person sits before a TV camera, the picture of his face is digitized and fed into the computer, a simple program extracts lines and locates a few feature points (the method is the same as that described in [32]) and, finally, his face is classified into one of seven categories, each of which is represented by a very famous person. Though the program was not very reliable, the attraction itself was very successful. A lot of people participated in it, a number of whose faces were stored on ten magnectic tapes. They include faces of young and old, males and females, with glasses and hats, and faces with a turn, tilt or inclination to a slight degree. It is of great interest to see how the program works with them. The rest of the data set comprises pictures I took for the purpose of investigating the performance and limitation of the program.

III-5-1. Summary of Results

The results are summarized in Table 3-1. The input faces were categorized into four groups:

(a) full face with no glasses or beard

(b) full face with glasses

(c) face with turn or tilt

(d) face with beard, and others

The first category corresponds to the faces which were assumed for the input of the program. The result of analysis was judged either "correct" or "incorrect" by human inspection. "Incorrect" includes erroneous results and unrecovered failures. In the case of incorrect results, the step in which the error or unrecovered failure took place is examined.

For pictures of the group(a), 608 pictures out of 670 were successfully analyzed, giving all of the face-feature points correctly. For about 200 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 3-1, the program works fairly well also on pictures of groups (b), (c) and (d) which do not satisfy the presumed constraints.

category of faces	number of faces	correct results	error or un-recovered failure	step in which the error or unrecovered failure occurred				
				face sides	nose mouth chin	chin contour	eyes	nose width
full face with no glasses or beard	670	608	62	5	14	17	19	7
full face with glasses	77	2	75	4	4	2	65	
face with turn or tilt	79	63	16	4	3	3	5	1
face with beard, and others	27		27		12	4		11

Table 3-1 Summary of results of analysis

III-5-2. Examples and Discussions

(1) Two sets of examples

Fig. 3-24 and Fig. 3-25 present two complete sets of examples. In each figure, part(a) is the original photograph and part(b) shows the gray-level representation of the digitized picture on the line-printer. Part(c) gives the binary picture in which the extracted feature points are indicated by dots. To verify how well the program located the feature points, their coordinates are marked on the enlarged original photograph as shown in part(d).

Fig. 3-26 shows the picture in which the same set of feature points is located by human. Comparison of Fig. 3-24(d) with Fig. 3-26 shows that the computer extraction is considerably good. Most results of pictures in the group(a) which are judged as correct are of this quality.

The locations of the feature points obtained in the binary picture may contain small errors: eye centers lie a bit on the outer side; face-side lines are apt to shift a little to the darker side. These phenomena stem from the property of the Laplacian operator used in the line extraction. The Laplacian operator, as is easily seen, produces a big positive output when it is on the darker side of the edge in the original picture. Thus, in the binary picture which is obtained by thresholding the differential picture, lines appear on the darker side of the real edge, rather than just on the right position.

In the next chapter, more precise locations are obtained by returning to the original picture, for the purpose of face identification.

(2) Face with facial lines or extra shadow

Facial lines and extra shadow in the face often result in troublesome line segments in the binary picture, hence failure or error of the analysis. In Fig. 3-27, the correct result is obtained because the search of the chin contour was limited to a small area based on the implicit model of the face. In the case of Fig. 3-28, the program was deceived by the several parallel line segments near the mouth and was not able to detect the face-side line.

(3) Face with glasses

In the case of faces with glasses, the program usually fails in the step of eye detection or gives erroneous eye positions, but this is only natural because the program was not taught that people may wear glasses!!; all the other results, however, are usually correct. Fig. 3-29 illustrates an example.

(4) Face with mustache

In the face with mustache, thick line segments appear between mouth and nose as shown in Fig. 3-30, and the nose (P and Q indicate the nose ends) is often located below the right position.

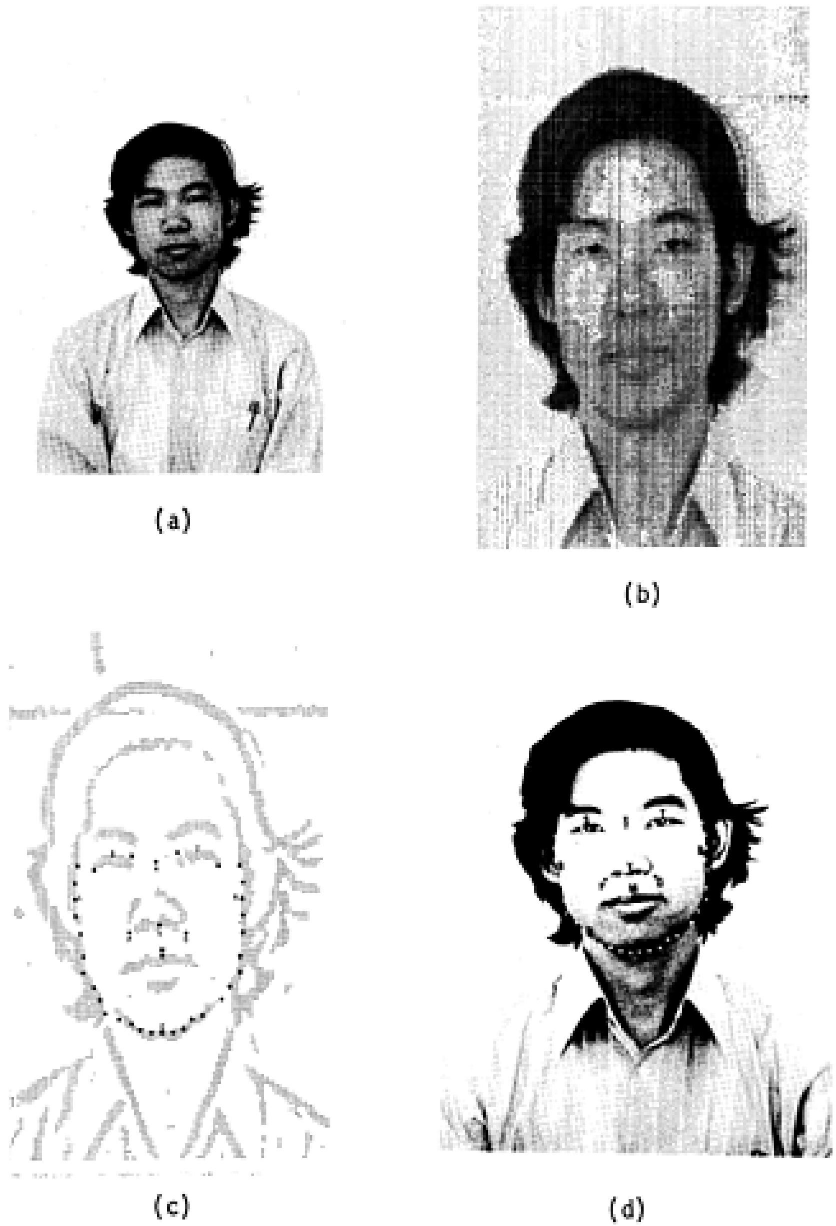


Figure 3-24 Result of the analysis : example 1.

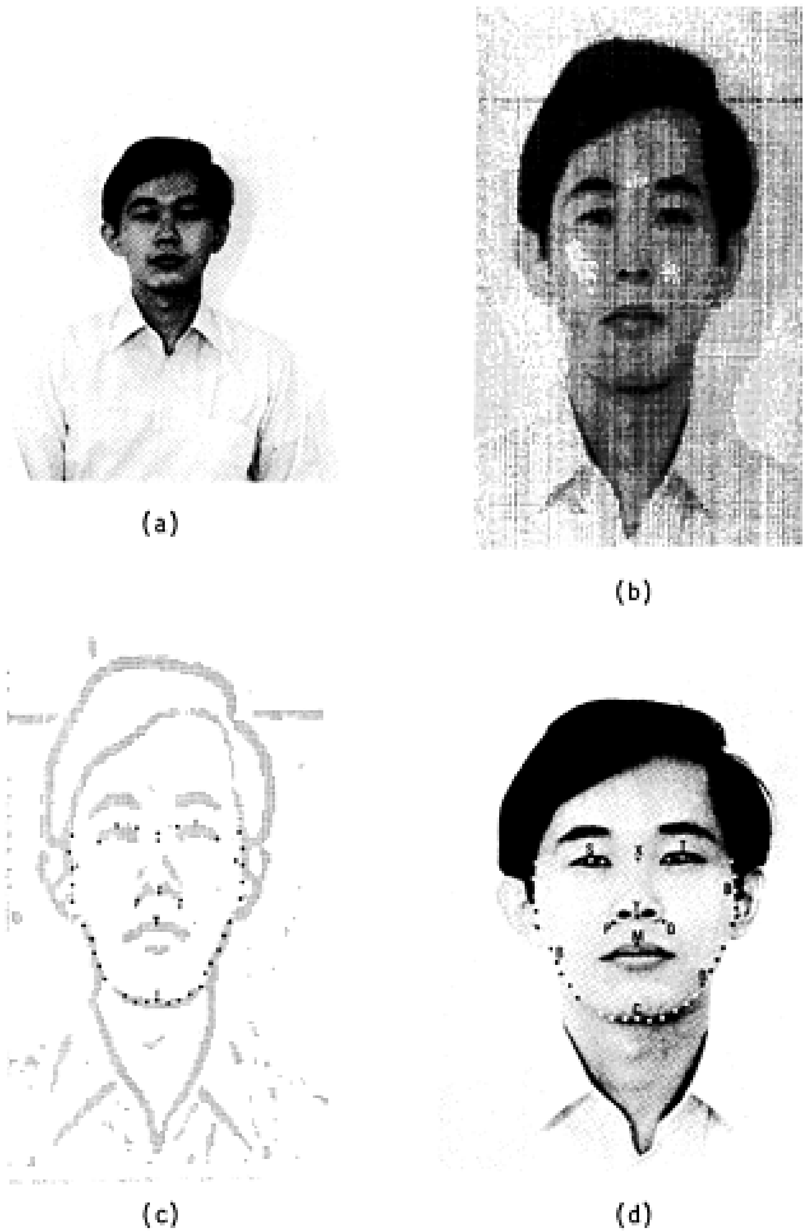


Figure 3-25 Result of the analysis : example 2.

Figure 3-26

Photograph with feature points located by human.

Figure 3-27 Face with facial lines : example 1.

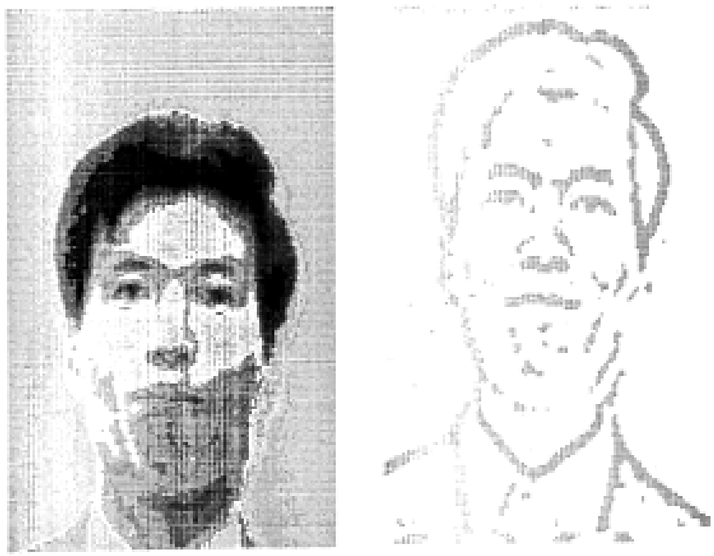


Figure 3-28 Face with facial lines : example 2.

Figure 3-29 Face with glasses.

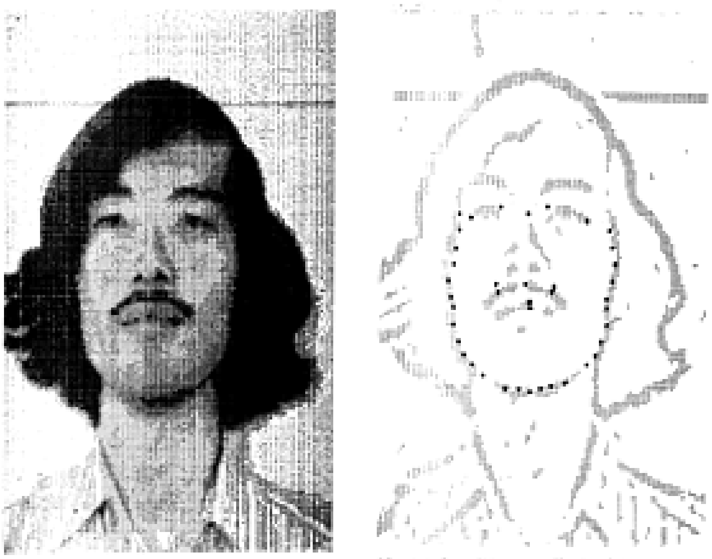


Figure 3-30 Face with mustache.

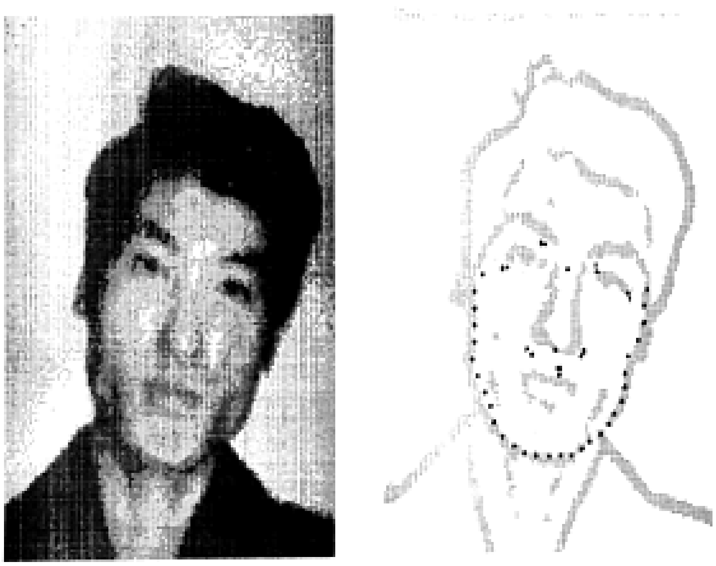


Figure 3-31 Face with turn or tilt : example 1.

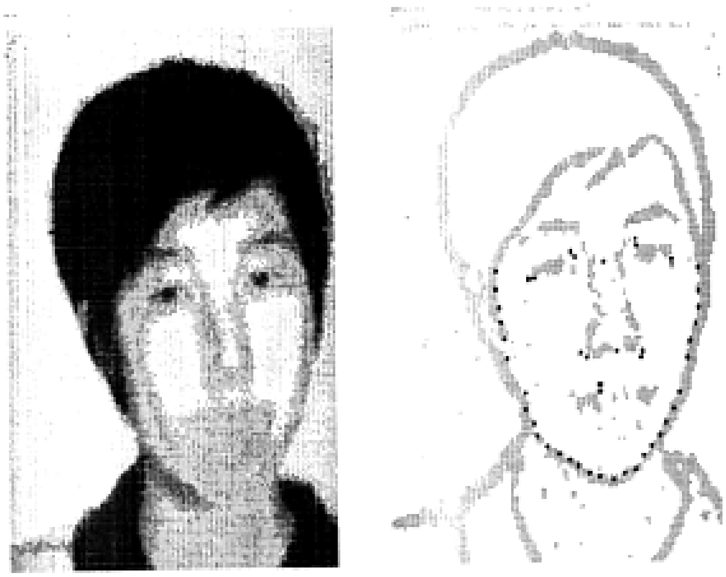


Figure 3-32 Face with turn or tilt : example 2.

Figure 3-33 Face with sparse chin-contour segments.

Figure 3-34 Noisy picture.

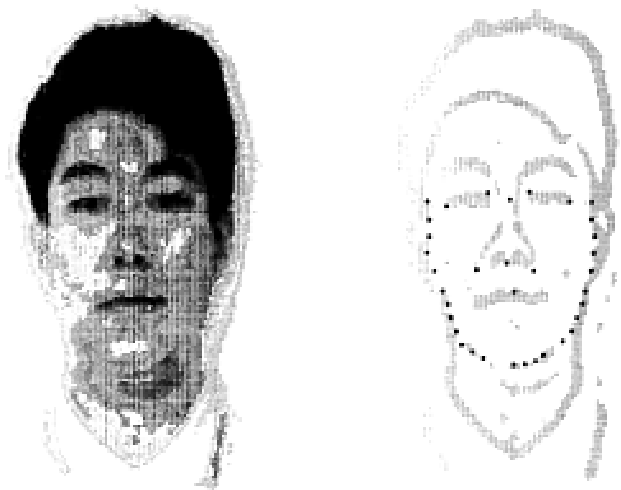
Figure 3-35

Picture of low contrast.

(a) Original photograph

(b) Digitized picture by the beam (00)

(c) Binary picture: the chin contour is not detected.



(a) Digitized picture of Fig. 3-35(a) by the beam (10)

(b) Result of the analysis

Figure 3-36 Effect of beam-intensity control

(5) Face with turn or tilt

Fig. 3-31 and Fig. 3-32 show the faces that involve a small extent of turn and tilt. Though the results are fairly good, the positions of mouth, nose ends and eye rectangle have some errors because no explicit compensation is made for the turn and tilt.

(6) Case of sparse chin-contour segments

Fig. 3-33 illustrates the case in which the feedback process with adaptive threshold setting in the chin-contour detection made it possible to detect a set of sparse line segments as forming the chin contour.

(7) Noisy pictures

A lot of noisy points exist in the picture of Fig. 3-34. However, they did not produce any effect on the analysis result, thanks to the use of integral projection and the predictive nature of the program. The operation of eliminating a small isolated segments beforehand may seem preferable, but sparse segments are often meaningfull as in the case of Fig. 3-33.

(8) Picture of low contrast

In the case of pictures of low contrast around the chin, there is no chin contour detected in the binary picture. Fig. 3-35 is a typical example. In this case, the beam-intensity control of the flying-spot scanner(see Fig. 2-6) displays its usefulness. The digital picture of Fig. 3-35(b) is sampled by the beam (00). The use of the beam (10) increases the contrast of the middle range in the digital picture as shown in Fig. 3-36(a). Therefore, in Fig. 3-36(b) the chin contour appears in the binary picture and the analysis program can detect it.

(9) Size variation

The program itself is able to accept the size variation of face up to 15 $\sim$ 20%. It may be readily extended so that larger variations are acceptable, by making combined use of the flexible scanning program of FSS (see II-5-1). Suppose a face in the input picture is too small. The program will fail in the first step of detecting the face sides L and R, but the size of the face can be estimated. Based on the estimation, finer scanning with a smaller spot size will probably produce a digital picture in which the face is of the acceptable size.

APPENDIX C contains supplementary examples. Much more examples would be necessary to completely show the excellent properties of the program, but the successful results of analysis of more than 800 pictures have demonstrated that the flexible picture analysis scheme with feedback which was employed in the program is very powerful for dealing with complex pictures.

CHAPTER IV

IDENTIFICATION OF HUMAN FACES

IV-1. Introduction

The recognition of a face is one of the most common human experiences, and yet it is very difficult to explain how this is done. Little research has been reported on machine identification of faces. This chapter describes an attempt at face identification of 20 people by computer.

Main emphasis is put on computer extraction of face features from the picture. The identification test of faces was conducted mainly to verify the results of the feature-finding algorithms, as well as to attempt an automatic visual identification of people. Therefore the method employed is rather standard and straightforward: the picture is first analyzed to locate feature points; then, facial parameters are calculated; a weighted Euclidian distance defined on these parameters is used to measure the similarity of faces.

IV-2. The Problems

To automate the identification of human faces, three problems arise as in other pattern-recognition tasks:

(1) Selection of effective features to identify faces

About face-features, a few works have been reported. Bledsoe[6] and Kaya and Kobayashi[16] used geometrical parametrization on the basis of coordinates of facial landmarks, whereas Goldstein, Harmon and Lesk[12] used subjective descriptors of features such as face shape, hair texture and lip thickness.

(2) Machine extraction of those features

Computer processing of human-face pictures for feature extraction has rarely been studied. Face-identification systems so far developed are man-machine system, in which feature extraction is performed more or less by human. In Kelly's work[17], only five measurements are drawn from the face, and it seems that measurements in the body picture, e.g., height and width of shoulders, play more important roles in his identification test.

(3) Decision-making based on the feature measurements

This is related to learning machines in pattern recognition. Many elaborate alogrithms to derive the optimal decision rule have been proposed. Actually, however, in such a complex problem as the identification of faces, the set of selected features mainly determines the performance of identification. In fact, Bledsoe and Kelly used a rather simple weighted Euclidian distance. In the tasks of suspect-face-file search and fingerprint-file search, it is more reliable and useful to find a small subset in which the target is surely included than to try to identify the best match in the file. The rank-ordering process devised in [12] provides quick reduction of target population.

At present the problems (1) and (3) are not of our major concern, though they are of interest. Computer extraction of facial features is our principal purpose. In this thesis, the geometrical parametrization was used to characterize a face; distances and angles among such points as eye corners, mouth extremities, nostrils, and chin top. These facial landmarks are located by computer picture processing. Decision-making uses a simple Euclidian distance defined on the feature parameters.

IV-3. Computer Extraction of Face Features

IV-3-1. Data Set

Machine identification of 20 people by their faces was attempted. They were all young people, 17 males and 3 females, without glasses, mustache or beard. Pictures of the subjects were taken in two series, i.e., two pictures for each person. The first and second pictures of the same person were taken in a different place with a time interval of one month. Thus, a collection of 40 photographs of a full face was used in the experiment. No special arrangements on lighting and other photographic conditions were made, except for asking the subject to turn a full face. The films were processed in a normal way, and positive prints with an appropriate size were produced to fit with the opaque-material head of the flying-spot scanner.

IV-3-2. Two-Stage Process

Computer picture processing is carried out to locate face-feature points in the face, such as eye corners, nostrils, chin contour and so on. As shown in Fig. 4-1, the whole process consists of two stages. The first stage is exactly the process described in CHAPTER III: the whole picture is scanned to produce a 5-bit digital picture of 140 x 208 elements; the sampling is made at every 3 addressable points by a 5 x 5 spot by means of the flexible scanning program of the flying-spot scanner; then the secondary differentiation and thresholding yield a binary picture, in which the picture analysis program locates a set of facial landmarks; an example of the processing result of the first stage is shown in Fig. 4-2.

The coarse scanning by a relatively large spot has eased the succeeding differential operation and feature-finding algorithms of the first stage, but, for the purpose of face-identification, the accuracy is insufficient. Once the locations of eyes, mouth, nose, etc. are known, at least approximately, one can return to the original picture to extract more accurate information by confining the processing to smaller regions and scanning with higher resolution. This refinement process is the second stage, which will be described in the following two subsections.

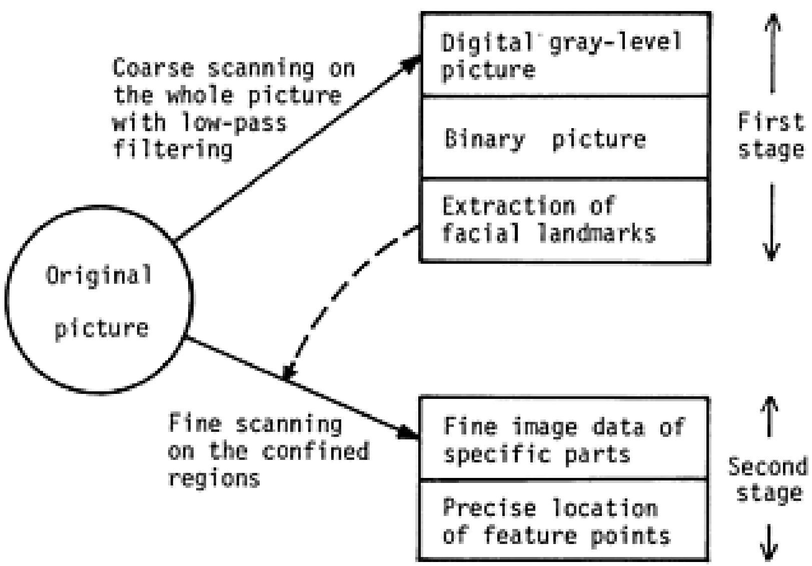


Figure 4-1 Two-stage process for computer measurement of features of human-face photographs.

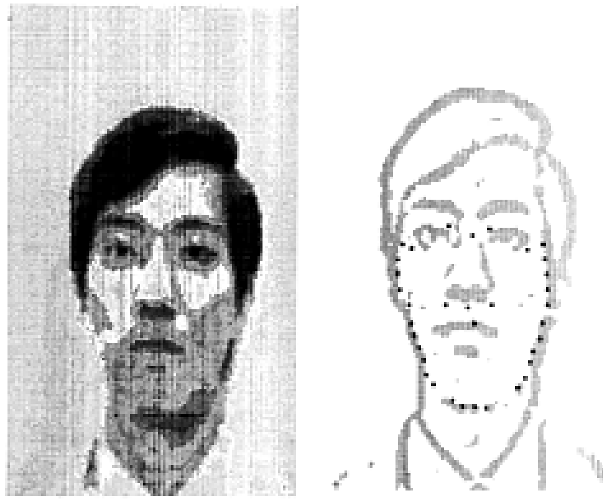


Figure 4-2 Result of the first stage of processing.

IV-3-3. Sampling of Finer Image Data

The first thing to do is to obtain finer image data of the confined portions. From the feature points obtained in the first stage, four small regions are settled as shown in Fig. 4-3. They correspond to the right eye, left eye, nose and mouth. This time, the scanning is made with the highest resolution, i.e., spot size = 1 x 1 (point spot) and sampling step = 1. The "best" beam intensity is selected to obtain the details. Four kinds of beam intensity are available as was explained in CHAPTER II (see Fig. 2-6). The best beam intensity to be used for each region is determined as follows. Take the nose region for instance. Fig. 4-4(a) shows the histogram of gray levels contained in the corresponding region of the digital picture which was scanned by the beam intensity (00) and used in the

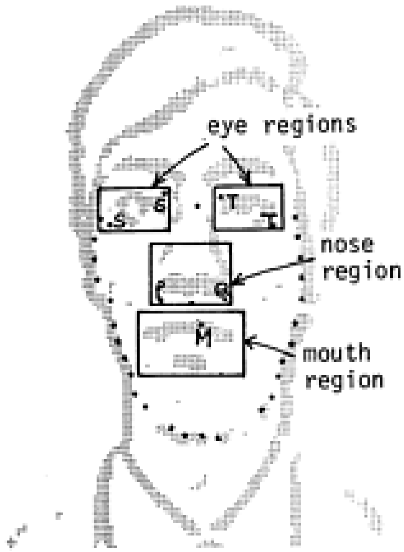


Figure 4-3

Four regions to be sampled in the second stage. They are determined by the positions of S, T, P, Q, and M.

Size of the regions:

Eye -- 70 x 40
Nose -- 80 x 50
Mouth -- 105 x 50 .

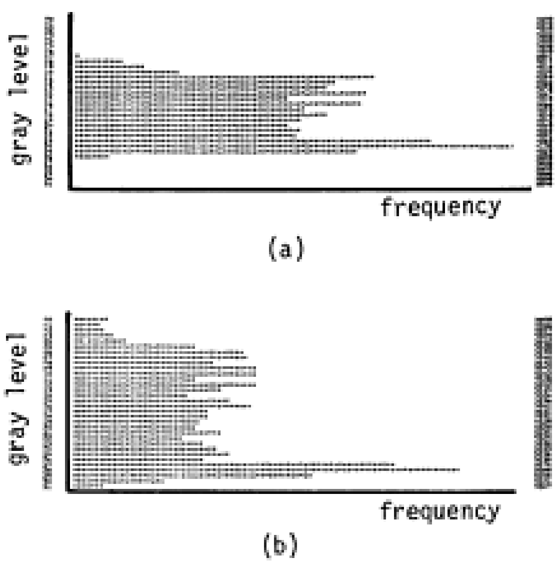


Figure 4-4 Histograms of gray levels.

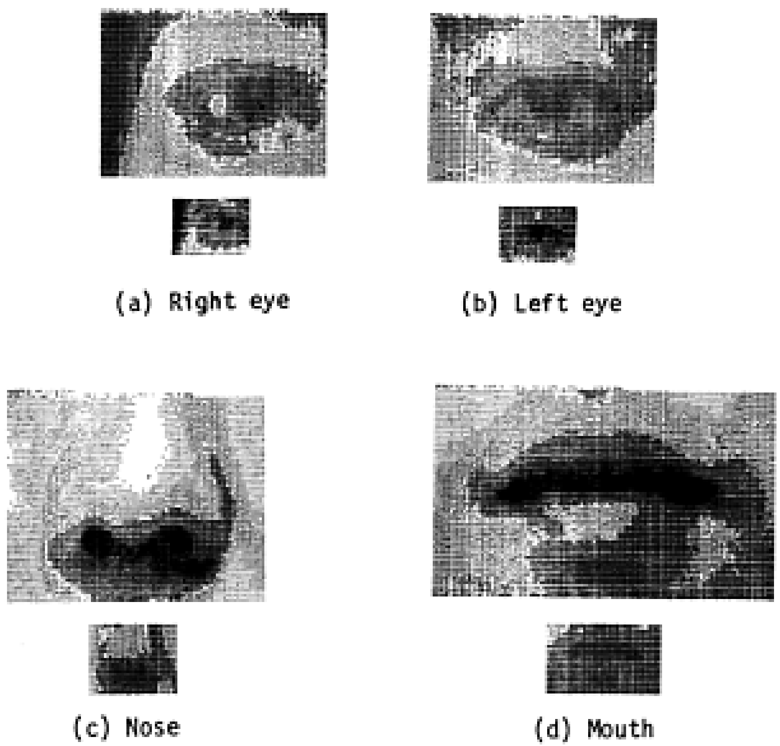


Figure 4-5 Printouts of detailed image data.

first stage (see Fig. 4-5(c)). The histogram indicates that the shades of gray in the region range from 8 to 26. Thus the beam intensity (10) will provide the best detail, because the middle range is digitized into 32 levels, and the darker and brighter ranges saturate to the lowest value 0 and the highest value 31, respectively. In fact, as shown in Fig. 4-4(b) the new histogram of the detailed image data of the nose region (Fig. 4-5(c)) proves the effect of this beam selection: the whole digital range is used effectively.

In this way, the best beam intensity is selected for each region: (00) for the region of wide-range intensity, (01) for dark region, (10) for medium, and (11) for bright. Fig. 4-5 shows the printouts of the detailed image data (upper) and those of the corresponding regions in the picture used in the first stage (lower). It can be seen that the upper pictures contain more details. This re-scanning process can be regarded as a kind of visual accommodation in both position and light intensity; the conditions of image-data sampling are adapted to the individual portions.

IV-3-4. Description of Extraction Processes

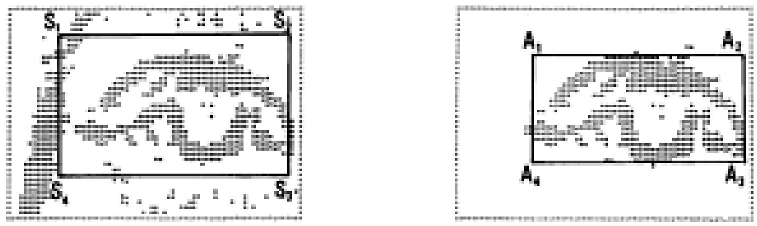
After acquisition of the finer image data, the next processing is to extract more precise information from each region. This time, all the operations used are very straight and simple: thresholding, differentiation, integral projection, and so on. Such simple operations suffice to get the facial details, because the program knows what part of the face it is looking at and what properties it deals with. Moreover it should be noted that this two-stage process could reduce the memory and time which might have been wasted for processing unnecessarily fine image data of irrelevant portions.

<u>(1) Eye</u>

Fig. 4-5(a) is the gray-level representation of image array in which the right eye is contained. For each picture element I_{ij}, the first derivative

$$dI_{ij} = |(I_{i-1j-1}+I_{i-1j}+I_{i-1j+1}) - (I_{i+1j-1}+I_{i+1j}+I_{i+1j+1})| + |(I_{i-1j-1}+I_{ij-1}+I_{i+1j-1}) - (I_{i-1j+1}+I_{ij+1}+I_{i+1j+1})| \quad (4\text{-}1)$$

is computed. Fig. 4-6(a) shows the binary picture in which the (ij) element has 1 if $dI_{ij}>\theta$ or 0 if $dI_{ij}<\theta$. The value of θ is set 15 in this experiment.



(a) Elimination of irrelevant portions

(b) Determination of eye rectangle

Figure 4-6 Detection of eye corners

The operations shown in Fig. 4-6 can locate the eye corners precisely. The first operation is to erase the connected components which cross the frame edge <u>and</u> lie outside of the circumscribed rectangle of the right eye $S_1S_2S_3S_4$ that has been determined in the first stage (Fig. 4-6(a)). This operation eliminates the portions that correspond to eyebrow, nose line, and hair. The circumscribed rectangle of the remaining components, $A_1A_2A_3A_4$ is determined (Fig. 4-6(b)); it indicates the precise location of the eye.

The same operation is performed on the left-eye region. The result is illustrated in Fig. 4-7; the four vertexes are displayed in the gray-level digital picture.

<u>(2) Nose</u>

In the nose region, the two nostrils are the features that can be consistently obtained, because they appear extremely dark.

Two small areas are fixed, in which the left and right nostrils may be contained, respectively. P-tile method is applied on each area to determine the threshold that singles out the nostril from the background: the threshold value is the minimum T such that the number of points which are darker than T exceeds a predetermined number P = 45. Points that are darker than T are located, and the fusion of distance 1 is performed on it. The center of gravity of the maximum component in each area is computed as the center of the nostril.

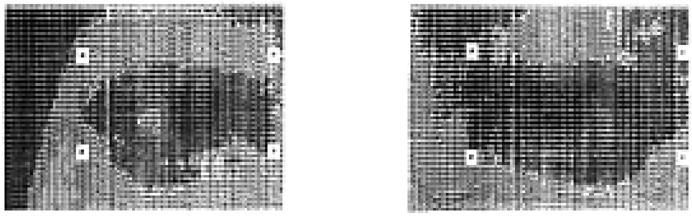
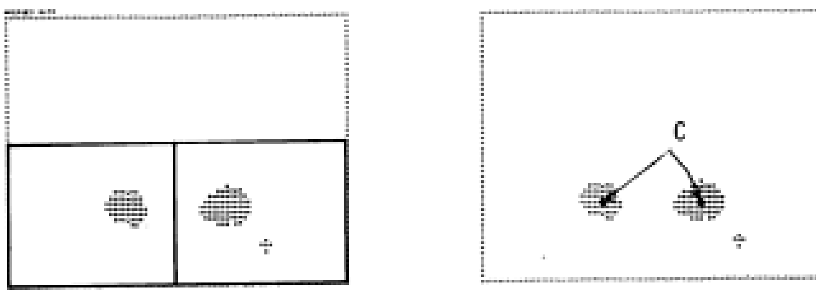


Figure 4-7 Eye corners.



(a) P-tile method and fusion

(b) Centers of gravity of maximum components

Figure 4-8 Detection of nostrils.

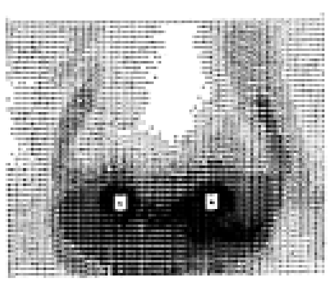


Figure 4-9 Centers of nostrils.

Fig. 4-8 illustrates this process. Fig. 4-9 shows the results: the centers of nostrils are displayed in the picture.

(3) Mouth

Lip thickness and the shape of mouths are very distinctive features of people, but it is not easy to extract them with consistency. Therefore, in the mouth region, only the position and breadth of the mouth are used.

The lips occupy a relatively dark and oblong area in the mouth region as can be seen in Fig. 4-5(d). As shown in Fig. 4-10, the vertical integral projection is first calculated. The deepest valley V corresponds to the dark portion where the upper and lower lips meet. It shows the vertical position of the mouth. Next, along this horizontal line, a narrow band is established as shown in Fig. 4-10. From the horizontal projection of this band, the breadth can be determined by detecting the darker part; the detail is explained in Fig. 4-10. The left and right extremities of the mouth located in this way are displayed in the picture of Fig. 4-11.

This completes the second stage of the processing. All the facial feature points which have been extracted in the first and second stages are used to calculate facial parameters.

Examples of the processing results are presented in APPENDIX D. They are a part of 40 pictures used in the identification experiment.

IV-4. Facial Parameters

A set of sixteen facial parameters $X = (x_1, x_2, \ldots, x_{16})$ shown in Fig. 4-12 is calculated. They all are ratios of distance and area, and angles to compensate for the varing size of pictures. In order to eliminate the differences of dimension and scale among the components of X, the vector X is normalized according to the following equation:

$$y_i = \frac{x_i - m_i}{\sigma_i} \qquad i = 1, 2, \ldots, 16 \qquad (4\text{-}2)$$

where $m_i = E[x_i]$, $\sigma_i^2 = E[(x_i - m_i)^2]$; and $E[\cdot]$ stands for the average

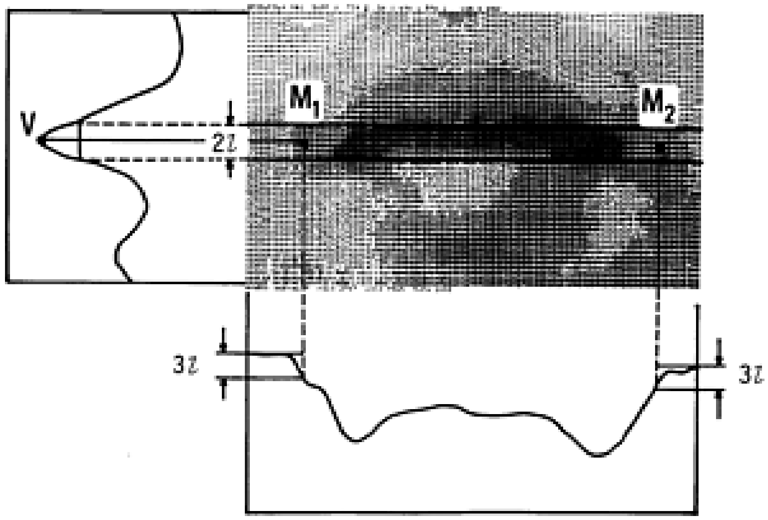


Figure 4-10 Determination of mouth extremities.

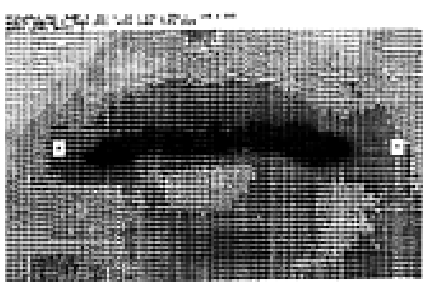
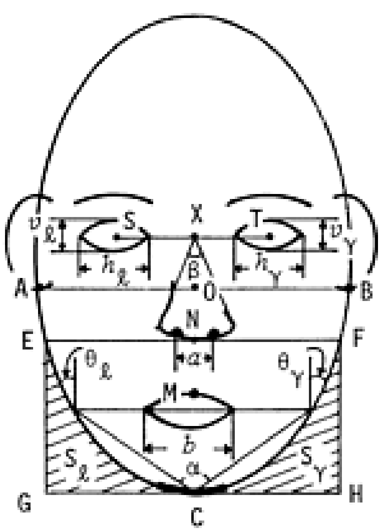


Figure 4-11

Mouth extremities.

over the whole set of data. That is, y_i means how far the parameter value x_i is from the mean $m_i = E[x_i]$, normalized by the standard deviation σ_i. y_i takes a positive value if x_i is greater than the mean and negative if smaller than the mean. Thus a picture is expressed by a vector $Y = (y_1, y_2, \ldots, y_{16})$.

On the other hand, the same set of facial parameters is measured by human hand in the enlarged photograph of the face using a millimeter scale and a transparent millimeter grid. The normalization is also performed. This is done to compare the feature measurements by computer with those by human. Hereafter, the superscript c or h is put on the parameter variable y_i when it is necessary to indicate whether it is obtained by computer or by human.



x_1 = AB/OC
x_2 = ST/AB
x_3 = NC/OC
x_4 = {curvature of the top of the chin
$x_5 = (\Delta EGC + \Delta FCH)/(S_\ell + S_r)$
$x_6 = \alpha$
$x_7 = \frac{1}{2}(\theta_\ell + \theta_r)$
$x_8 = \beta$
x_9 = XN/XC
x_{10} = NM/XC
$x_{11} = (h_\ell + h_r)/ST$
$x_{12} = (v_\ell + v_r)/ST$
$x_{13} = a/ST$
$x_{14} = b/ST$
x_{15} = ST/XN
$x_{16} = (h_\ell + h_r)/(v_\ell + v_r)$

Figure 4-12 Facial parameters.

Before going to identification experiment, the effectiveness or repeatability of parameter values has been examined in the following way. Pick up a picture ${}_{l}Y$. Let ${}_{l}Y'$ be the other picture of the same person. Then

$$ {}_{l}U_{i} = |{}_{l}y_{i} - {}_{l}y'_{i}| \tag{4-3} $$

is hopefully the minimum of the set

$$ {}_{l}S_{i} = \{|{}_{l}y_{i} - {}_{k}y_{i}|, \; l,k = 1, 2, \ldots, 40, \; k \neq l\}, \tag{4-4} $$

but this is not always the case. Let ${}_{l}U_{i}$ be the ${}_{l}r_{i}$-th smallest of ${}_{l}S_{i}$. Table 4-1 tabulates the average α_{i} and standard deviation β_{i} of ${}_{l}r_{i}$ over the whole collection of pictures, i.e.,

$$ \alpha_{i} = E[{}_{l}r_{i}] $$
$$ \beta_{i} = \sqrt{E[({}_{l}r_{i} - \alpha_{i})^2]} \quad . \tag{4-5} $$

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	Average
α_i^c	12.2	15.6	16.9	16.2	19.1	16.3	15.0	13.7	13.7	17.1	11.8	10.7	12.7	17.3	10.5	10.5	14.3
β_i^c	9.7	9.6	9.5	10.1	10.4	11.3	10.8	11.0	10.3	9.9	7.7	7.5	8.8	10.9	8.6	8.5	
α_i^h	10.2	10.1	14.8	13.8	17.9	12.0	11.9	19.3	12.2	9.3	15.3	11.7	11.4	9.0	13.6	10.5	12.7
β_i^h	7.1	8.8	10.0	10.5	8.8	8.2	9.8	9.7	8.5	7.0	10.5	8.9	9.4	7.9	10.0	8.1	

Table 4-1 Repeatability of parameter values.

The upper and lower rows correspond to, respectively, measurements by computer and by hand. It can be said from the table that the reliability of individual parameters is not very high. The accuracy of computer extraction is slightly inferior to that of human extraction.

IV-5. Identification Test

The phase of identification test of faces was run separately from the phase of picture processing. The whole collection of 40 pictures was processed, and all the parameter measurements were obtained both by computer and by hand. The resultant normalized vectors $_{l}Y^{c}$ and $_{l}Y^{h}$ for $l = 1 \sim 40$ were recorded on a magnetic tape.

The experiment was run as follows. The collection of 40 pictures was divided into two sets of equal size. The first set that comprises 20 pictures, one picture for each person, is the reference set R, or the set of known individuals. The remaining 20 pictures are used as the test set T or the set of unknown individuals. Now the problem is: given a picture in the test set T, find the picture of the same person in the reference set R.

We chose a simple distance

$$d(Y, Y') = \sum_{i=1}^{16} |y_i - y_i'|^k \qquad (4\text{-}6)$$

as a measure of similarity between the pictures Y and Y'. Therefore,

Feature measurement \ k	1/2	1	2	3
Computer	10	9	10	10
Human	15	13	13	12

Table 4-2 The number of people who were correctly identified when all the parameters were used. A better performance is obtained when ineffective parameters are omitted in decision-making.

Person \ k	Rank of the real target picture: 1/2	1	2	3
A	1	1	1	1
B	1	1	1	1
C	2	1	1	1
D	3	3	1	1
E	1	1	1	1
F	4	5	5	5
G	1	2	3	6
H	2	2	2	1
I	1	1	1	2
J	2	1	1	1
K	1	1	1	1
L	2	2	2	5
M	1	1	1	1
N	2	1	1	1
O	2	1	1	1
P	1	1	1	1
Q	5	5	2	1
R	1	1	1	1
S	2	2	1	1
T	3	2	2	2
Average	1.9	1.8	1.5	1.8
Number of correct identification	9	12	14	15

Table 4-3

Results of identification test; feature measurement by computer.

Person \ k	Rank of the real target picture: 1/2	1	2	3
A	1	1	1	1
B	2	2	2	2
C	1	1	1	1
D	2	1	1	1
E	2	1	1	1
F	1	1	1	1
G	1	1	1	1
H	6	3	3	5
I	1	1	1	1
J	1	1	1	1
K	1	2	2	5
L	1	1	1	1
M	5	5	5	5
N	1	1	2	2
O	1	1	1	1
P	1	1	1	1
Q	4	3	2	3
R	1	1	1	1
S	1	1	1	1
T	1	1	1	1
Average	1.8	1.5	1.5	1.8
Number of correct identification	14	15	14	14

Table 4-4

Results of identification test; feature measurement by human.

when given a picture $Y \in T$ in the test set, the answer is the picture $\tilde{Y}$ in the reference set such that

$$\min_{Y' \in R} d(Y, Y') = d(Y, \tilde{Y}). \qquad (4\text{-}7)$$

The identification test was carried out for $K = \frac{1}{2}$, 1, 2, 3. Table 4-2 summarizes the results. The correct identification is 45 ~ 75%.

Better results are obtained when ineffective parameters are not used in calculating the distance. The parameters, the sum of whose α_i and β_i is large, are regarded as ineffective.* In the computer extraction, y_5, y_6 and y_{14} are discarded and in the human extraction, y_5, y_8 and y_{11}. Therefore only thirteen parameters are used. The results of identification test are presented in Table 4-3 (computer extraction) and Table 4-4 (human extraction). The entry number stands for the rank of the real target picture: "1" means that the correct individual is identified. In both tables, 15 out of 20 people are correctly identified in the best cases.

From this limited experiment it is difficult to judge whether a fully automatic identification of faces is feasible or not, but the experiment has shown that the computer is certainly capable of extracting facial features that can be used for face identification of people, almost as reliably as a person.

---

*This criterion may be justified as follows. Suppose ${}_i r_i$ follows a normal distribution $N(\alpha_i, \beta_i)$. Then $C_i = \alpha_i + \beta_i$ is the rank below which 84% of cases fall.

CHAPTER V

CONCLUSION

V-1. Summary of This Thesis

This thesis has been devoted to the descriptions of a new picture processing system, and computer analysis and identification of human faces. Principal results obtained in this research are summarized as follows:

(1) In CHAPTER II, a new flexible picture processing system has been developed. It employed system-subsystem organization; the mini-computer and graphical I/O devices form a satellite subsystem of the host computer. Intimate intercomputer communications are realized between the two computers by means of a special software support. The form of this computer-computer connection has shown one of the new directions to the inhouse computer network, in which high-speed transmission of a large volume of data and message switching of short commands co-exist.

(2) In addition, the satellite can perform intelligent jobs for itself. It provides several characteristic facilities for picture processing such as highly flexible scanning and film-recording programs. Many examples of usage of the facilities have demonstrated their wide and potential applicability.

(3) In CHAPTER III, the analysis of human-face photographs has been performed as a task of pictorial information processing. Given

a photograph of the face, the analysis program can locate facial feature points, such as eye corners, nose, mouth, chin contour and so on. More than 800 pictures were successfully processed by the program.

(4) The flexible picture analysis scheme was presented and employed in the analysis program of human face. It consists of a collection of rather simple subroutines, each of which works on the specific part of the picture, and elaborate combination of them with feedback interwoven makes the whole process flexible and adaptive. The face-analysis program developed in this thesis is one of the first ones that implement this type of analysis scheme with feedback for picture analysis. Excellent properties of the program as well as successful results of the analysis have proved that this manner of organizing programs is very promising to the analysis of complex pictures.

(5) In CHAPTER IV, visual identification of people by their faces was tested. 15 out of 20 people were correctly identified relying solely on measurements obtained by a computer. This is the first success of machine identification of human-face photographs.

(6) Though simple, the two-stage process employed in extracting precise features has a significant meaning. It is an example of distant feedback: high-level decision affects and controls input sampling stage; the facilities that the new picture processing system provides were demonstrated to be indispensable.

V-2. Picture Structure and Analysis Method

This section provides a discussion on the generalized concept of the method used for processing human-face photographs.

Several methods of pattern recognition so far developed are first discussed briefly from the point of manipulating the picture structure.

The techniques of correlation in two dimensions can be used for pattern recognition. Prepared a file of standard objects, a given scene is correlated with each member of the file and it is concluded that it most resembles that object which gives the maximum correlation. This type of approach is often termed "template matching" and it has an advantage that high-speed computation of correlation may be carried out by means of optical devices. Although straightforward in concept, the template-matching method, we know, is less potent for recognizing patterns with large variations, because it is sensitive to translation, magnification, brightness, contrast and orientation.

Statistical pattern recognition is another important approach. The pattern is treated as a multi-dimensional vector in the feature space, the components of which denote individual observations, e.g., the area of black portions in a character. A decision rule that divides the feature space into subspaces assigns an unknown pattern to one of predefined classes. Many algorithms have been proposed for constructing the optimal decision rules. But in this method, as is often pointed out, the choice of features is more crucial than the decision rule.

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[18], Narasimhan[24], Ledley[19], and others took the so-called linguistic approach. As shown in Fig.5-1, 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.

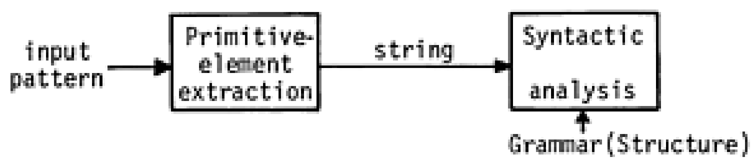


Figure 5-1 Introduction of structure in the liguistic approach; this scheme is not satisfactory for complex pictures.

The linguistic approach presents certain difficulties in its application. In addition to the difficulty in describing a two-dimensional image in the string form, the method suffers from the fact that the stage of primitive-element extraction does not know what element is meaningful and what is noise because it is separated from the stage of syntactic analysis. Another reason for this is that primitive elements are usually line segments and arcs: they are neither subparts nor substructures of the picture, but only convenient units of processing. Recently a method appeared which uses a description of the picture in terms of properties of regions (e.g. circular) and relations between them (e.g. adjacent), instead of lines[2]. It may have the same kind of disadvantages unless the relational structures of objects which are stored do not direct the process of pictorial analysis of a given scene.

In order to overcome this difficulty, Shaw[33] proposed a top-down goal-directed syntax analyzer. It "parses" the picture in the analogous manner to a classical top-down string parser: the grammar is explicitly used to direct the analysis and to control the calls on picture processing routines for primitive elements. Unfortunately the relation the analyzer uses is only the head-and-tail concatenation, thus limiting its applicability.

Now, let's discuss the significance of the method I employed in analyzing pictures of human faces. Some difficulties would be encountered if the linguistic approach were taken for human-face analysis. Suppose that small line segments, arcs and the like are selected as the primitives. Then the resultant rules for composing a face would be intolerably complicated and awkward. Moreover, a lot of noisy line segments unrelated to the face structure, perhaps from wrinkles and hair style, would be annoying. Thus it is more convenient to consider a face to be composed of larger substructures such as eyes, nose, mouth and so on, having the relatively macro-positional relations. Recall that the eyes, for instance, are recognized as such only in the context of a face; in fact, many other components show the shape of the eye in the picture. This means that the search for the component with the shape of the eye should be done in the region in which the eye is expected to exist.

The scheme of Fig. 5-2 which I employed is better for this purpose: first the region in which a particular part (e.g. head top) is predicted to exist is determined by taking full advantage of *a priori* knowledge about the object class of pictures 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 under given constraints. Each procedure can take advantage of the fact that it knows what to look for and where to look at. For rough and tentative estimation of

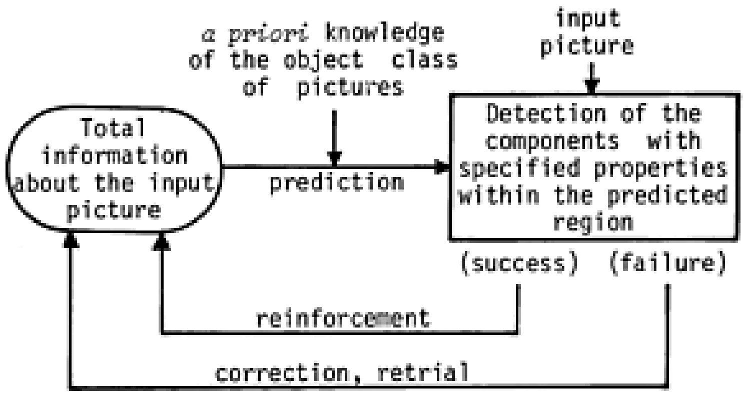


Figure 5-2 Generalized concept of the picture analysis scheme employed in this thesis.

search areas, procedures insensitive to noise are used which might not be so accurate, but once the search areas for specific components are established, accurate procedures are employed which determine the exact properties of the components, and which sometimes incorporate *ad hoc* methods. If the program fails to detect the component 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 the whole process very flexible. Note that it also simplifies each step of analysis: perfect results need not to be obtained in the once-and-for-all processing, for the errors, when detected at later steps, can be recovered.

The feedback may involve the input scanning of the picture to obtain more proper image data. A simple example is the two-stage process described in the first part of CHAPTER IV for extracting precise features of the face. This "distant" feedback prevents waste of memory and time for treating unnecessarily fine image data.

The scheme described above resembles somewhat the top-down goal-directed syntax analyzer of Shaw[33] in that both employ "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, 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.

V-3. Picture Analysis and Problem Solving

The picture analysis scheme used in the analysis of human-face photographs is fundamentally as follows: a collection of routines is prepared, the member of which works on the specified region and yields certain results when given a set of data it needs; the routines are called into action in the appropriate sequence with retrial and feed-back interwoven.

This scheme, when advanced, leads to the view that a picture-analysis program should be a problem-solving program: that is, the analysis does not proceed in a predetermined order, but the program decides the best sequence of analysis steps depending on individual input pictures, on the basis of *a priori* knowledge and evidence so far gathered about the given input. The program selects from a collection of routines one that is expected to give the most reliable result. The selected routine is executed and the total information about the input increases. In this way, the program reaches the goal of describing the given picture. 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 is the most essential in this method, and one possible approach is discussed by Nagao[21].

In the present version of the face-analysis program, the correction, retrial and backtracking are not sufficient. The main reasons for this are: (1) the face is not an object that requires such a high-level capability of problem solving, and (2) the program coded in the assembler language makes it difficult to implement decision making with high complexity, recursive call of routines and storage of variable-length data.

The second point will become a serious obstacle to manage a very complex picture. Therefore in the advanced picture-analysis program the processing of low-level image data will be coded, for instance, in the assembler language for high-speed operation and the part of higher-level decision making will be programmed in the symbol manipulation language such as LISP to facilitate the usage of problem-solving techniques. The system described in CHAPTER II is designed so that it can be used to develop such advanced picture-analysis programs.

The problem of computer picture processing has many different phases. Sometimes, a single transformation performed upon the whole picture may suffice the purposes, among them uniform enhancement or smoothing. When it comes to analysis or understanding of the picture, the problem will involve much deeper aspects; structure, knowledge, and problem solving. The results of this thesis suggest that, rather than individual techniques of picture processing, it is more essential to combine them in a proper manner based on the context and structure of the picture. The work reported here will serve to advance picture processing in that direction.

REFERENCES

[1] R.G.Barr:"POLLY1: Operator-assisted bubble chamber film measuring system", *Review of Scientific Instruments*, Vol.39, No.10, Oct. 1968.

[2] H.G.Barrow and R.J.Popplestone:"Relational Descriptions in Picture Processing", in *Machine Intelligence* 6, Michie(ed.), pp. 377-395.

[3] P.L.Bastien and L.A.Dunn:"Global Transformations in Pattern Recognition of Bubble Chamber Photographs", *IEEE Trans.*, Vol. C-20, No.9, pp.995-1001, Sep. 1971.

[4] F.Beck:" The Special Features of POLLY: How They are Used for Automatic Scanning", *IEEE Trans.*, Vol.C-20, No.9, p.1002, Sept. 1971.

[5] L.A.Belady, M.W.Blasgen, C.J.Evangelisti and R.D.Tennison:" A Computer Graphics System for Block Diagram Problems", *IBM Syst. J.*, No.2, pp.143-161, 1971.

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

[7] H.Chan and W.W,Bledsoe:" A Man-Machine Facial Recognition System; Some Preliminary Results ", Panoramic Research Inc. Palo Alto, Cal., May 1965.

[8] M.A.Fischler and R.A.Elschlager:" The Representation and Matching of Pictorial Structures ",Lockheed Palo Alto Research Lab. Lockheed Missiles & Space Company, Inc. Cal., LMSC-D243781, Sept. 1971.

[9] J.D.Foley:" An Approach to the Optimum Design of Computer Graphic Systems", *C.ACM*, Vol.14, No.6, pp.380-390, June 1971.

[10] K.Fukushima and Y.Yamaguchi:" Picture I/O Device of Digital Scanning Type ", *Proc. 2nd Image Conf. Japan*, pp.17-20, 1971 (in Japanese)

[11] A.J.Goldstein, L.D.Harmon and A.B.Lesk:" Identification of Human Faces ", *Proc. IEEE*, Vol.59, p.748, 1971.

[12] A.J.Goldstein, L.D.Harmon and A.B.Lesk:" Man-Machine Interaction in Human-Face Identification ", *Bell Sys. Tech. J.*, Vol.51, No.2, pp.399-427, Feb. 1972.

[13] S.B.Gray:"Technical Description of the GRAFIX I Image Processing System", Information International, Los Angels, Calif. USA.

[14] L.D.Harmon:"Some Aspects of Recognition of Human Faces", in *Pattern Recognition in Biological and Technical Systems*, O.J.Gruesser(ed.), New York Springer.

[15] N.M.Herbst and P.M.Will:" An Experimental Laboratory for Pattern Recognition and Signal Processing ", *C.ACM*, Vol.15, No.4, pp.231-244, April 1972.

[16] Y.Kaya and K.Kobayashi:"A Basic Study on Human Face Recognition", in *Frontiers of Pattern Recognition*, S.Watanabe(ed.), p.265, 1972.

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

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

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

[20] R.S.Ledley, L.S.Rotolo, M.Belson, J.Jacobsen, J.B.Wilson and T.Golab:"Pattern Recognition Studies in the Biomedical Sciences", *Proc. SJCC*,pp.411-430, 1966.

[21] M.Nagao:"Picture Recognition and Data Structure", in *Graphic Languages*, p.48, North-Holland, 1972.

[22] M.Nagao and T.Kanade:"Edge and Line Extraction in Pattern Recognition", *Proc. Inst. Elect. Comm. Engrs. Japan*, Vol.55, No.12, pp.1618-1627, Dec. 1972 (in Japanese).

[23] G.Nagy:"State of the Art in Pattern Recognition", *Proc. IEEE*, Vol.56, No.5, pp.836-862, May 1968.

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

[25] D.A.O'Handley, E.S.Beckenbach, K.R.Castleman, R.H.Selzer and R.J.Wall:"Picture Analysis Applied to Biomedicine", presented at the US-Japan Seminar on Picture and Scene Analysis, Kyoto, Japan, July 1973.

[26] K.K.Pingle and J.M.Tenenbaum:"An Accommodating Edge Follower", *Proc. AI Conf.*, London, Session No.1, 1971.

[27] I.Pless:"The PEPR System", *IEEE Trans.*, Vol.NS-12, No.8, p.279, Aug. 1965.

[28] L.G.Robertz:"Machine Perception of Three-dimensional Solids", in *Optical and Electro-Optical Information Processing*, pp.159-197, Cambridge, Mass. MIT Press, 1965.

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

[30] T.Sakai, T.Kanade, M.Nagao, K.Tabata, S.Kitazawa, K.Ohnishi and Y.Ohta:"Inhouse Computer Network for Information Processing and Some Applications", Presented at the US-Japan Seminar on Computer Assisted Chemical Research Design, Hawaii, July, 1973.

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

[32] T.Sakai, M.Nagao and M.Kidode:"Processing of Multi-Leveled Pictures by Computer —— The Case of Photographs of Human Face ——", *Jour. Inst. Elect. Comm. Engrs, Japan, Abstracts*, Vol.54, No.6, p.27, June 1971.

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

[34] P.G.Stein, L.E.Lipkin and H.M.Shapiro:"Spectre II: General-Purpose Microscope Input for a Computer", *Science*, Vol.166, pp.328-333, Oct. 1969.

LIST OF PUBLICATIONS AND TECHNICAL REPORTS

1. " Picture Structure and Processing Method — the Case of Human-Face Photographs — "; T.Sakai, M.Nagao and T.Kanade, *Proc. Joint Conf. of Elect. Engrs. Japan*, p.820, Oct. 1971(in Japanese)

2. " On a Self-adptive Pattern Classifier Employing an Iterative Algorithm "; T.Sakai and T.Kanade, *Trans. Inst. Electronics Comm. Engrs. Japan Abstracts*, Vol.54, No.11, p.20, Nov. 1971 (Originally appeared in Japanese; *Trans. IECEJ*, 54-C 11, pp.979-986).

3. " Picture Processing System with a Minicomputer-controlled Input Device and a General-Purpose Computer "; T.Sakai, M.Nagao and T.Kanade, Technical Report of the Professional Group on Automaton and Language of IECEJ, AL-72-47, July 1972 (in Japanese)

4. " Computer Analysis and Classification of Photographs of Human Faces "; T.Sakai, M.Nagao and T.Kanade, *Proc. First USA-JAPAN Computer Conference*, pp.55-62, Oct. 1972.

5. " Edge and Line Extraction in Pattern Recognition "; M.Nagao and T.Kanade, *Proc. Inst. Electronics Comm. Engrs. Japan*, Vol.55, No.12,pp.1618-1627, Dec. 1972 (in Japanese).

6. " Computer Analysis of Photographs of Human Faces "; T.Sakai, M. Nagao and T.Kanade, *Trans. Inst. Electronics Comm. Engrs. Japan Abstracts*, Vol.56, No.4, p.29, April 1973 (Originally appeared in Japanese; *Trans. IECEJ*, 56-D, 4, pp.226-233).

7. " Inhouse Computer Network for Information Processing and Some Applications "; T.Sakai, T.Kanade, M.Nagao, K.Tabata, S.Kitazawa, H.Ohnishi and Y.Ohta, presented at the Seminar on Computer Assisted Chemical Research Design, July 2-6 1973, Hawaii USA.

8. " Picture Processing System by Computer Complex "; T.Sakai, T. Kanade, M.Nagao and Y.Ohta, presented at the USA-JAPAN Seminar on Picture and Scene Analysis, July 23-27 1973, Kyoto Japan.

9. " An Experiment on Visual Identification of Faces by Computer "; T.Sakai, M.Nagao and T.Kanade, presented at the USA-JAPAN Seminar on Picture and Scene Analysis, July 23-27 1973, Kyoto Japan.

10. " Picture Processing System Using a Computer Complex "; T. Sakai, T. Kanade, M. Nagao and Y. Ohta, *Computer Graphics and Image Processing*, to appear.

APPENDIX A

DESCRIPTION OF HARDWARE AND SOFTWARE OF GIRLS

A-1. GIRLS-Interface Controller

This section describes in detail the functional capabilities of the GIRLS-interface controller(Fig. A-1), which contains special interfaces designed for the picture processing system. There are four connections; MACC-NEAC, MACC-FSS, MACC-CRT and MACC-Graf/pen.

A-1-1. NEAC-MACC Connection

As shown schematically in Fig. A-2, the GIRLS interface performs three functions; data transfer, interrupt and status sense which take place between NEAC 2200/200 and MACC 7/F. The two computers are connected by a set of 64 coaxial cables. Table A-1(a) shows the instructions of NEAC and MACC related to this connection.

1) data transfer

The data flow and registers in MACC-NEAC connection are shown in Fig. A-3. Data transfer is done in 6 bit-parallel mode. One character (6 bits) of NEAC corresponds to the lower 6 bits of one word of MACC: as illustrated in Fig. A-2, seventh through sixteenth bits of MACC word are neglected when transferred from MACC to NEAC, and 0's

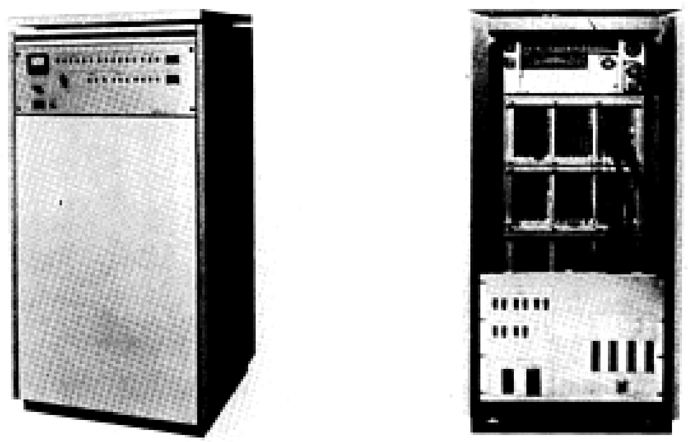


Figure A-1 GIRLS-interface controller.

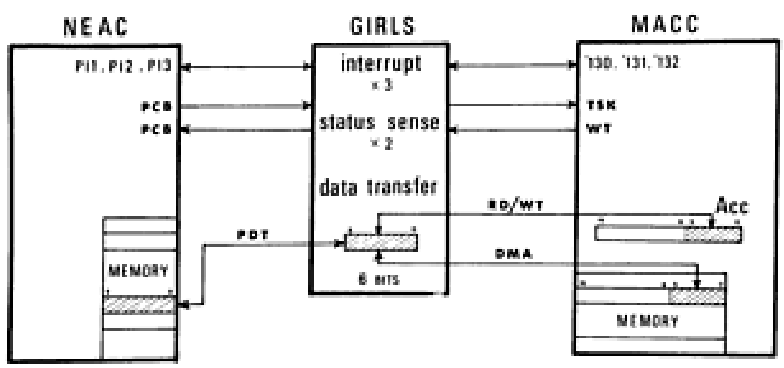


Figure A-2 Schematic diagram of MACC-NEAC connection.

NEAC 2200/200

DATA I/O	POT A,C_1,C_2,C_3		A: Starting address of buffer C_1: Channel number C_2: Trunk number C_3: Variant(see Fig. A-4)
CONTROL	PCB $*,C_1,C_2,C_3$		
	C_3	02	Reset PA3 (PA3=0)
		03	Set PA3 (PA3=1)
		04	Reset PA5 (PA5=0)
		05	Set PA5 (PA5=1)
		20	Quit POT
		21	Interrupt MACC 7/F (level 2)
		22	Set PEC (PEC=1)
		23	Resume POT
		51,52,53	Mask WT/'C2, WT/'C3, RECORD-MARK interruption, respectively
		61,62,63	Unmask WT/'C2, WT/'C3, RECORD-MARK interruption, respectively
		70	Mask all the interruption
		71	Unmask all the interruption
		74	Reset all the interruption flip-flops
TEST	PCB A,C_1,C_2,C_3		
	C_3	10	If either a POT is in execution or POT interruption to MACC is masked, then jump to A, else go to the next
		30	If PS0=1 then jump to A, else go to the next
		31	If PS1=1, then jump to A, else go to the next
		41,42,43	If, respectively, WT/'C2, WT/'C3, or RECORD-MARK interruption is set, then jump to A, else go to the next

MACC 7/F

	opcode	device address	machine code (hexadecimal)	function
DATA I/O	WT	C1	E1C1	Write lower 6 bits of ACC in MN-register and request NEAC to read them in
	WT	C8	E1C8	Write 16 bits of ACC in MRWC register
	RD	C0	E0C0	Read NM register into ACC(lower 6 bits)
	RD	C1	E0C1	Read NM register into ACC(lower 6 bits) and request NEAC to transfer the next 6 bits
	RD	C2	E0C2	Read C register into ACC(lower 6 bits)
CONTROL	WT	C0	E1C0	Quit POT
	WT	C2	E1C2	Interrupt NEAC via PI1
	WT	C3	E1C3	Interrupt NEAC via PI2
	WT	C4	E1C4	Set PS0 (PS0=1)
	WT	C5	E1C5	Reset PS0 (PS0=0)
	WT	C6	E1C6	Set PS1 (PS1=1)
	WT	C7	E1C7	Reset PS1 (PS1=0)
	WT	C9	E1C9	Set PEC (PEC=1)
	MSF	CA	E9CA	Mask PCB interruption
	MSF	CB	E9CB	Unmask PCB interruption
	MSF	CC	E9CC	Mask RECORD-MARK interruption
	MSF	CD	E9CD	Unmask RECORD-MARK interruption
	MSF	CE	E9CE	Mask POT interruption
	MSF	CF	E9CF	Unmask POT interruption and resume POT
TEST	TSK	C0	EBC0	Skip if POT is in execution
	TSK	C1	EBC1	Skip if NEAC is ready to transfer one character
	TSK	C2	EBC2	Skip if PA3=1
	TSK	C3	EBC3	Skip if PA5=1

Table A-1(a) Instructions for MACC-NEAC connection.

opcode	device address	machine code (hexadecimal)	function
WT	D0	E1D0	INITIALIZE AND RESET ; reset to multi-level mode
WT	D1	E1D1	WRITE COMMAND : enable FSS
WT	D2	E1D2	SET X : write lower 10 bits of ACC in GX register
WT	D3	E1D3	SET Y : write lower 10 bits of ACC in GY register
WT	D4	E1D4	WRITE STROBE : turn on the beam and move the spot to (X,Y)
WT	D5	E1D5	SET BEAM INTENSITY : select one of the four types of digitization
WT	D6	E1D6	SET SLICE LEVEL AND CHANGE TO BINARY MODE
WT	FE	E1FE	ADVANCE FILM BY ONE SPROCKET
WT	FF	E1FF	REVERSE FILM BY ONE SPROCKET
RD	D4	E0D4	READ DATA : read the digitized value (5 bits in multi-level mode or 1 bit in binary mode)
TSK	D0	EBD0	SCANNER READY : skip if FSS is ready
TSK	F0	EBF0	FILM TRANSPORT AUTO : skip if film transport is auto mode
TSK	F1	EBF1	FILM TRANSPORT READY : skip if film transport is not ready

Table A-1(b) Instructions for FSS control.

opcode	device address	machine code (hexadecimal)	function
RD	E0	E0E0	READ X COUNTER
RD	E1	E0E1	READ Y COUNTER
MSF	E0	E9E0	Unmask DATA-READY interruption
MSF	E1	E9E1	Mask DATA-READY interruption

Table A-1(c) Instructions for Graf/pen control.

opcode	device address	machine code (hexadecimal)	function
WT	E7	E1E7	Write lower 10 bits of ACC in CX register
WT	E8	E1E8	Write lower 10 bits of ACC in CY register
WT	E9	E1E9	Turn on the beam if the lowest bit of ACC is 1
WT	EA	E1EA	Change to WRITE-THROUGH operation
WT	EB	E1EB	Change to NON-STORE operation
WT	EC	E1EC	Reset to STORE operation
WT	ED	E1ED	Erase the stored image
WT	EE	E1EE	Change to VIEW mode
WT	EF	E1EF	Return to HOLDING mode

Table A-1(d) Instructions for CRT control.

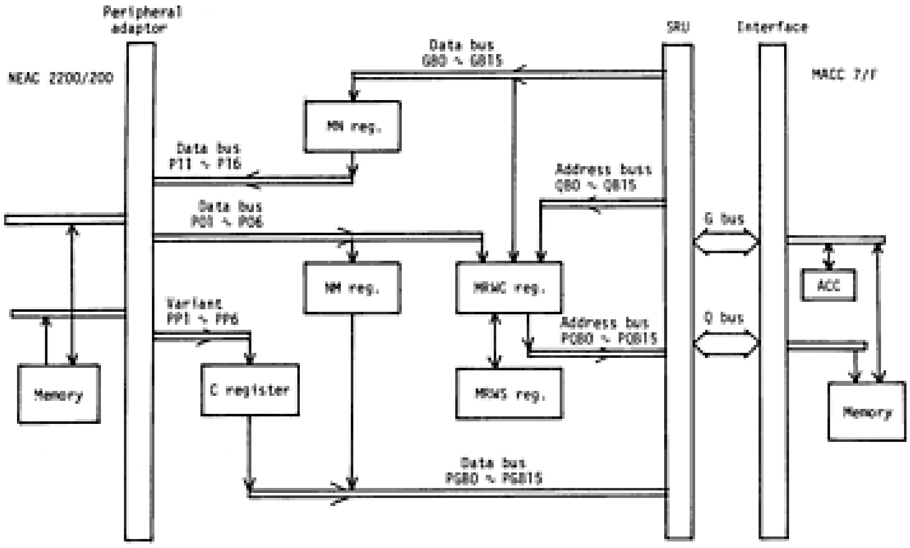


Figure A-3 Data flow and registers in NEAC-MACC connection.

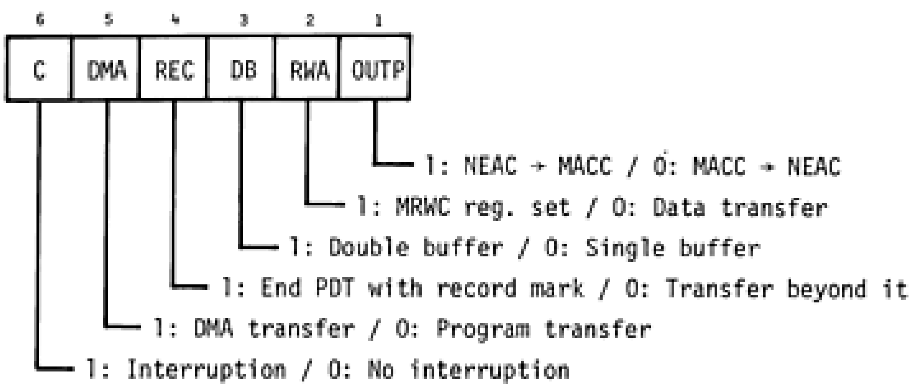


Figure A-4 C_3 of PDT instruction.

are set there when NEAC to MACC. Bit serial transmission, which is often employed in data-communication networks, might be another method. In our system, however, the situations are different. The two computers employ completely different bit systems. NEAC is 6-bit character-oriented and MACC is 16-bit word-oriented. They have stronger connection than mere data-communication. So the information to be transferred includes raw pictorial data, object machine code and high-level command. Hence, data editing and decomposition are to be carried out depending on the individual cases by software.

There are two kinds of method for data transfer.

(i) program transfer

The content of the accumulator of MACC is transmitted by a WT (output) or RD(input) instruction in the program. The transfer rate depends on the program, and the maximum rate is 12 μsec/6 bits.

(ii) DMA(Direct Memory Access) transfer

The data is transmitted directly from/to the memory of MACC. The rate is 6 ∿ 12 μsec/6 bits or in average 750 kbps. This high transmission rate permits quick transfer of a large volume of data which is very common in picture processing.

In both cases, NEAC executes a PDT instruction before the transfer of one block, and then the data flow between NEAC memory and the register of GIRLS interface.

A PDT instruction has the following format:

PDT A, C_1, C_2, C_3.

"A" is the address of the first location of the memory area to be transferred. C_1 and C_2 indicate the channel and trunk numbers, respectively. C_3 is a 6-bit control character to specify I/O conditions. Each bit was given a specific meaning as shown in Fig. A-4. Various combinations realize a variety of data-transfer functions. The value of C_3 is set to a special register(C register) in GIRLS interface just when the PDT instruction starts. Since MACC can read the C register, the program in MACC is able to know the I/O condition specified by the PDT, thus to react properly.

DB bit of C_3 has a special function. It enables double buffers to be used alternatively. To illustrate it, suppose that double buffers, A and B in NEAC and P and Q in MACC, are prepared as shown in Fig. A-5, and that data are transferred from MACC to NEAC by means of DMA (thus DMA=1, OUTP=0). DB=1 affects as if the next location of B_n were A_1, and the next of Q_n were P_1. Therefore, the transfer from P buffer to A buffer and the transfer from Q to B occur alternatively without resetting of the start locations of buffer areas by the program. Thus MACC program can write in P buffer while Q buffer is transferred to NEAC, and write in Q while P is transferred. In this way, the DB function enables continuous high-speed data aquisition and transfer to occur at the same time.

Among many other functions, temporary or complete stoppage of data transfer can be made automatically or by a program command from NEAC or MACC, in order to permit flexible data transfer between the two computers.

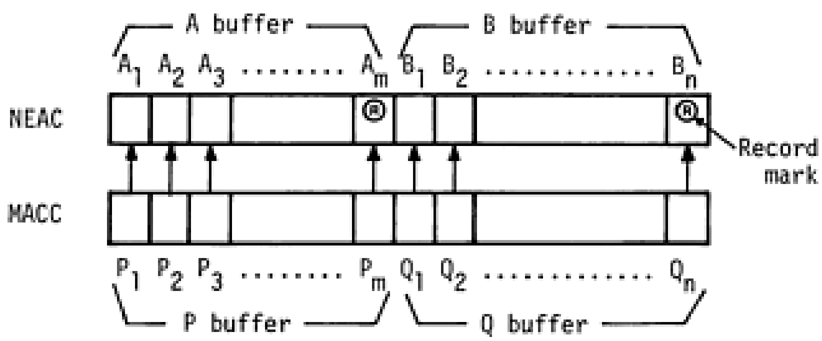


Figure A-5 Data transfer by means of double buffers.

2) Interruption

(i) interruption to NEAC (1 level, 3 sources)

Two program interruptions caused by WT instructions of MACC.

Record mark interruption caused when the character with a record mark is accessed during PDT execution. It indicates the end of transfer of one block bounded by the record mark.

(ii) interruption to MACC (3 levels)

Program interruption caused by a PCB instruction of NEAC.

Record mark interruption during PDT execution.

PDT interruption caused when PDT instruction whose C bit of C_3 is 1 starts. In case this interruption occurs, actual data transfer suspends, and begins only after the RESUME command is executed in MACC. Therefore, after a MACC program checks C_3 of the PDT by reading C register in GIRLS interface, it can postpone the actual data transfer, or even cancel the PDT, depending on its own program status.

3) Status Sense

Two status-sense lines for each direction NEAC↔MACC(PSQ, PS1, PA3, PA5). A program of one side can sense and know how far the program of the other has gone.

A-1-2. MACC-FSS Connection

All the sampling is done in a random scanning mode under the MACC program control, for, as explained in II-2, necessary scanning area is not always the whole frame, but scan of the specific area under the specific condition is desirable. Fig. A-6 is a functional diagram of the FSS controller. Table A-1(b) lists up the control instructions of FSS.

First, the X and Y coordinates of the scanning point are specified to 10 bits each and set to GX and GY registers, respectively. Then, a WRITE STROBE command is sent. It turns on the beam, and transfers the coordinate from GX and GY registers to X and Y registers, whose outputs are then fed into two digital-to-analog converters. The light produced at the programmed X-Y location on CRT is divided into two beams by a half mirror. One beam goes upward, and is focused onto the film plane. The light transmitted through the film enters the film photomultiplier whose output forms A channel. The other beam passes straight to the opaque material plane. The light reflected by the material is collected by four photomultipliers with a mirror tube. The summation of those outputs (2 Vp-p signal) is B channel. By manual selection, one of three analog channels, A, B, A-B, is fed into

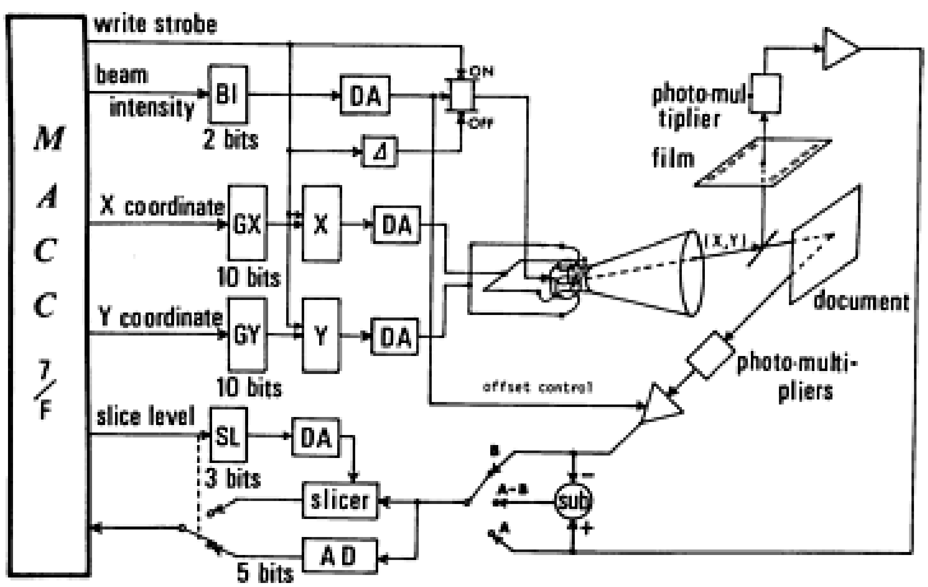


Figure A-6 Functional diagram of FSS controller.

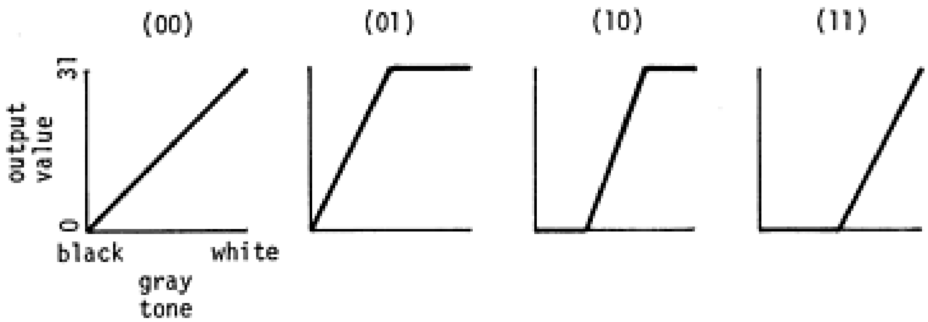


Figure A-7 Relations between gray tone and output value for each beam intensity.

both an analog-to-digital converter and the slicer, the threshold of which is a programmably settable number of 3-bit-precision. Finally, the five-bit-precision output of A/D converter or the binary output of the slicer is read into the accumulator by a RD DATA command. After the WRITE STROBE command, the RD DATA command has to wait for the time of moving the spot (max 11 μsec), stabilization of electro-optical circuits (2 μsec) and analog-to-digital conversion (2 μsec), but in this while the computer can do other computation, typically to determine the next scanning point, and thus wastes no time. In this way, the sampling of one point can be done within 15 μsec.

If the time interval between the successive WRITE STROBE commands is less than 95 μsec, the beam continues to be on, but if the next does not come before 95 μsec, then the beam is turned off to protect the phosphor screen of the CRT.

Among a number of program settable parameters, the "beam intensity" is worth mentioning. The two-bit number selects one of four relationships between digitized output and brightness of the material to be sampled (see Fig. A-7).

This function permits image-data sampling in the most desirable condition depending on the dynamic range and contrast of each material.

For the beam (00), 32 quantization levels are spread uniformly over the whole range of brightness, whereas for the beam (01), they are concentrated into the darker range, and the brighter range gives saturated output 31. The beams (10) and (11) correspond to the middle and brighter ranges, respectively. The brightness digitization, therefore, has substantially 6-bit resolution for gray scale. This control is accomplished by changing the second grid voltage of the CRT in two levels (slope of I/O relation) and the offset voltage of the video preamplifier (window of the digitization). This range and contrast control as well as other functions plays an important role in the "flexible scanning program" which has been described in II-5-1.

The film transport advances or reverses 35 mm perforated film by the sprocket wheel connected to a stepping motor. Frame changing time is about half a second. Fig. A-8 is a picture of the film transport.

The FSS is also used as a film recorder to record computer outputs on raw film, readily by replacing a scanner head with a camera head. The control is almost in the same manner as input, with only exception that the time of "beam on" is program-controlled in order to produce shades of gray on the film.

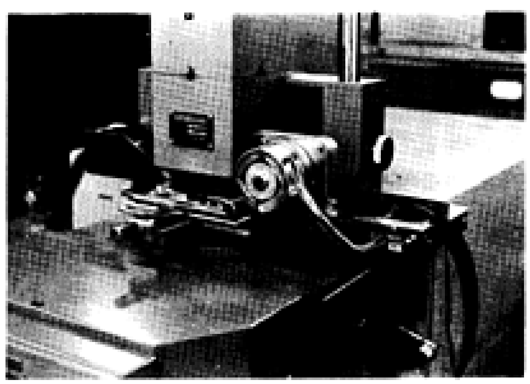


Figure A-8 Film transport.

A-1-3. MACC-Graf/pen Connection

Fig. A-9 shows the block diagram for the Graf/pen and interface, and the related instructions of MACC 7/F are listed in Table A-1(c). The spark is initiated by pressure of the stylus on the tablet. X and Y counters start with the spark trigger. Soundwaves from the spark propagate through the air to the sensors and cause the counters to stop: the counters then contain 10 bit numbers proportional to the X and Y distances from the stylus spark to the sensors. After these numbers are settled in the counter, a "data ready" signal is generated and it causes interruption to MACC. Then the MACC program

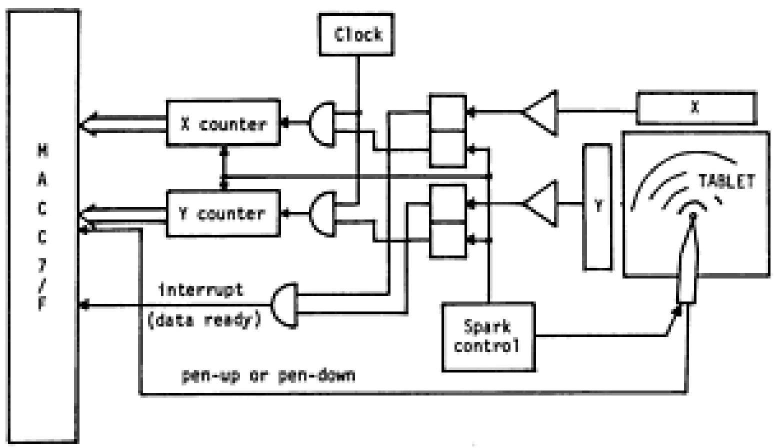


Figure A-9 Diagram of Graf/pen controller.

reads in the X-Y coordinate by RD commands. The data format is given in Fig. A-10. Since coordinate pairs are generated repeatedly at as fast as 200 pairs/sec, or in every 5 msec. at maximum, MACC affords to engage in additional processing on the data, e.g. smoothing, interpolation etc., while reading data intermittently.

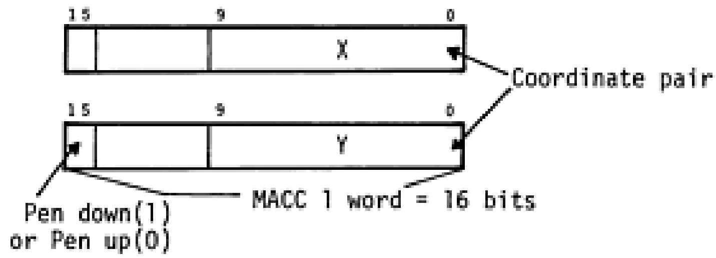


Figure A-10 Data format of Graf/pen.

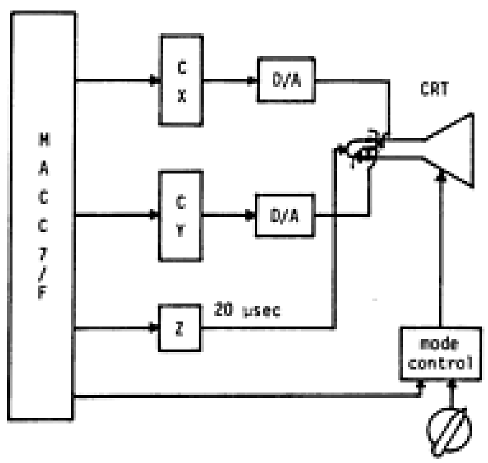


Figure A-11 Diagram of CRT controller.

A-1-4. MACC-CRT Connection

The logic and interface for the storage CRT is very simple. The diagram of Fig. A-11 and instructions of Table A-1(d) explain almost everything. The writing beam's position is specified by 10-bit-precision coordinates. There are three operating modes; STORE (stored display), NON-STORE (just like convensional CRT), and WRITE-THROUGH (illuminate the writing beam's location without storage, in addition to the stored display). Choice of the mode is made by a program command or a manual switch on the control unit. In the STORE-mode of operation, the writing beam has to stay for 20 μsec at the writing position to store the data, and the deflection time is about 5 μsec + 3.5 μsec/cm. It follows that the maximum storage rate is about 40k points/sec.

A-2 Software Support of System

The GIRLS system has been constructed so that more efficient, systematic use of the graphical I/O devices under the control of MACC 7/F may be realized by means of computer-computer communication between NEAC 2200/200 and MACC 7/F. To this end, a special software support has been developed. This section gives a technical description of the software.

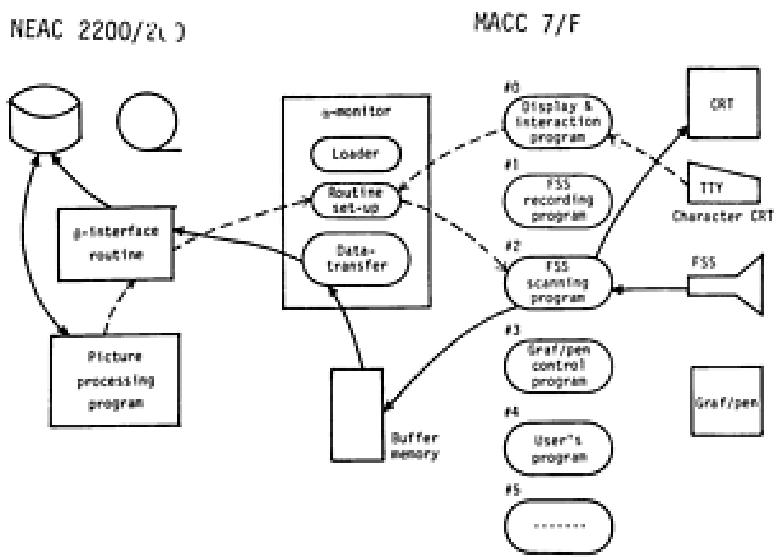


Figure A-12 Software organization of GIRLS.

A-2-1. Organization

Fig. A-12 shows a schematic diagram of software organization and data flow in solving a picture processing problem. It consists of four parts;

(1) User's problem-oriented program in NEAC:
It analyses the picture and manages the main sequence of analysis procedure. For instance, the part for manipulation of image data may be written in assembler language, and the part for decision-making may be written in LISP.
(2) A collection of routines for versatile class of operations in MACC:
It includes scanning, preprocessing, display, man-machine interaction, and feature measurements.
(3) α-monitor in MACC
(4) β-interface routine

(3) and (4) support intercomputer communication; they handle the demands for a program set-up from NEAC to MACC and the demands for a file access from MACC to NEAC, and control the data transfer between the two computers. Under the control of α-monitor and β-interface routine, the data transfer is extremely simplified by macro-instructions that enable MACC users to use the units of disk and magnetic tape of NEAC in exactly the same way as though they were attached to MACC.

A-2-2. α-monitor

α-monitor resides in the main memory of MACC 7/F and controls execution of programs. Because of the limitations imposed by a mini-computer, its major function, unlike a large-scale operating system, is just to facilitate intercomputer communication. It accepts a command from NEAC to cause an appropriate program in MACC to run, and manages data transfer.

All the programs under the control of α-monitor are identified by the intrinsic program number. It consists of four hexadecimal digits as shown in Fig. A-13; the first three are package number and the last one is unit number. A unit of program is an independent routine that has one specific function, e.g., FSS scanning. As many as one package (16 units) of programs are able to exist in the main memory at a time, as long as the memory space allows. The registration of a program is made just by writing its entry address in a particular table.

1) Set-up of program run

There are three ways for setting up a program to run in MACC 7/F.

(i) A program in NEAC is able to call routines registered in MACC by sending via β-interface routine a calling segment of the format shown in Fig. A-14; the first three characters (which correspond one word of MACC) indicate the program number to be set up; then follow parameters or arguments of the program, the size of which is up to 31 x 3 characters. The user is responsible for preparing the segment with appropriate program number and arguments. It is just like a subroutine call in FORTRAN: CALL PROG#(ARG1, ARG2, ..., ARGn).

(ii) An operator at MACC 7/F presses the console-interrupt button to communicate with α-monitor. Then he can enter the program number and arguments via TTY or PTR, which causes the corresponding routine to run.

(iii) A program in MACC can call other routines by executing the macro-instruction CALL NEXT. About macro-instructions, see 4).

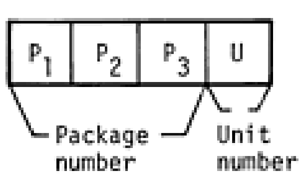


Figure A-13

Program number.

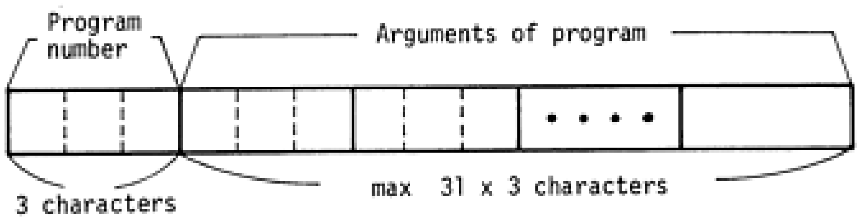


Figure A-14 Calling segment from NEAC to MACC.

2) End of routine run

A routine in MACC quits to run and is closed by executing a STOP macro-instruction. Then the control returns to α-monitor. It sends an end message to β-interface routine to notify the end of routine run.

3) Macro-instructions for data transfer from/to NEAC

These provide MACC users with facilities for data transfer from/to the core memory, disk, and MT of NEAC. Available instructions are listed in Table A-2(a). By using them MACC can share the storage devices and lineprinter with NEAC, which offers ease of programming and economical advantage. In summary, communication between programs in two computers is accomplished by two methods; (1) command message by sending a calling segment, and (2) a large volume of data by sharing storage devices.

4) Macro-instructions for calling α-monitor

To help communication between α-monitor and a user's program, several macro-instructions have been prepared. Table A-2(b) shows the names and functions.

Name	Format	Function
WRITE CORE	JS /I/ '10; 6; <buffer address>;	Transfer one word of MACC to the lower 16 bits of 3 characters of NEAC. The location of the buffer in NEAC is specified by a RESET macro of β-interface routine.
SET DISK	JS /I/ '10; 7; CC; TT; RR;	Set CC(cylinder), TT(track), and RR(record) of the disk device.
READ DISK	JS /I/ '10; 8; <buffer address>;	Read two records of the disk(256 characters) into the <buffer>.
WRITE DISK	JS /I/ '10; 9; <buffer address>;	Write 256 characters of the <buffer> into two records of the disk.
REWIND MT	JS /I/ '10; 10; <MT no.>;	Rewind <MT no.>.
FORWARD MT	JS /I/ '10; 11; <MT no.>;	Forwardspace <MT no.>.
BACKSPACE MT	JS /I/ '10; 12; <MT no.>;	Backspace <MT no.>.
READ MT	JS /I/ '10; 13; <MT no.>; <buffer address>;	Read one record (256 characters) from <MT no.> to the <buffer>.
WRITE MT	JS /I/ '10; 14; <MT no.>; <buffer address>;	Write 256 characters of <buffer> in <MT no.> as one record.
SET B&T	JS /I/ '10; 15; <device no.>;	<device no.>= <MT no.> or '20(disk) Usage: SET B&T T/I/'24 IM/'24 } any times RESET B&T
RESET B&T	JS /I/ '10; 16;	A user can, seemingly, store any length of data in the <device> without paying any attention to the location and size of the buffer in MACC : α-monitor prepares a fixed buffer, the last location of which is memory-protected; a user writes data in this buffer successively by the repetition of " T/I/'24;IM/'24; "; the violation of memory protection tells α-monitor that the buffer becomes full; it transfers data to the <device> of NEAC and resets the content of '24 to the first location of the buffer, so that the whole process may repeat.

Table A-2(a) Macro-instructions for data transfer from/to NEAC.

Name	Format	Function
STOP	J /I/ '14;	Quit the execution of a user's program, return the control to α-monitor and notify β-interface routine of the end of the program run.
CALL NEXT	JS /I/ '15; <program no.>;	Start the execution of the program specified by <program no.>.
CALL MACRO	JS /I/ '10; <macro no.>; <argument>; < .. >;	Start the execution of the macro routine specified by <macro no.>. The function is just the same as JS/I/*+1;EXT/macro;.
SET EXIT	JS /I/ '11; <interrupt no.>; <entry address>;	Connect the exit routine of <interrupt no.> with <entry address> so that when that interruption occurs the program starts to run.
RESET EXIT	JS /I/ '12; <interrupt no.>;	Make the interruption of <interrupt no.> unallowable.

Table A-2(b) Macro-instructions for calling α-monitor.

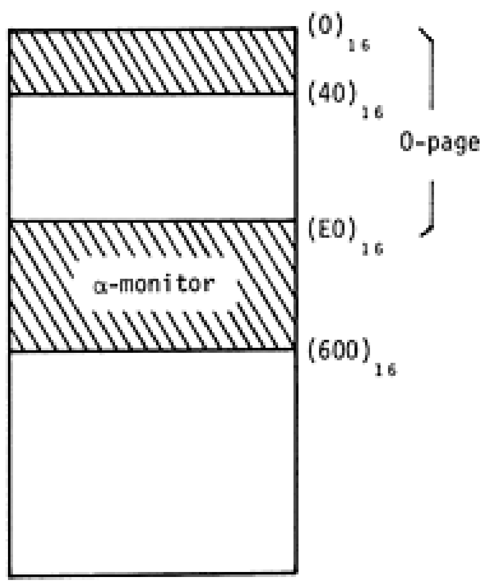


Figure A-15

Core map of MACC 7/F.

5) Loader

In case the program requested to be set up is not in the main memory of MACC, the loader rolls it in from a magnetic tape of NEAC or a small disk of MACC. Fortunately in MACC 7/F, if carefully coded, a unit of program can be relocatable in the level of machine language. All the facility programs of GIRLS are coded in such a way.

In addition to the above functions, α-monitor contains I/O macro-instructions for TTY and PTR, the routine for power fault, etc. α-monitor is divided into modules and so its size depends on their combination; the minimum organization including the loader occupies about 1.3 kwords of core. The core map of this case is illustrated in Fig. A-15. α-monitor itself consumes as many as 64 words in the 0-page (address 0 ∿ 255), for it is desirable that as large 0-page area as possible is open to users for writing efficient programs.

A-2-3. β-Interface Routine

β-interface routine sends a calling segment to MACC in order to request α-monitor for setting up the corresponding routine to run. It responds to file-access demands from MACC to make the peripherals available to MACC users. β-interface routine runs mainly in the external mode of NEAC in response to the interruptions from MACC, and therefore other programs can run in the normal mode.

There are two macro-instructions that β-interface routine handles;

RESET AND CALL MACC
RESET

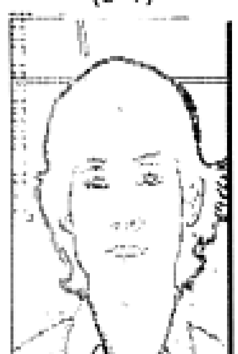
The first is used when NEAC user calls routines in MACC, whereas the second is used when the routine set-up is to be done in the MACC side.

COMPARISON OF LINE-DETECTION OPERATORS

(a)

(b-1)

(b-2)



(c-1)

(c-2)

(d-1)

(d-2)

Example 1

(a) Gray-level picture

(b) Laplacian operator (Fig. 3-8)

(b-1) θ = 25
(b-2) θ = 35

(c) Robertz operator (Fig. 3-10(a))

(c-1) θ = 4
(c-2) θ = 2

(d) Maximum of differences (Fig. 3-10(b))
(d-1) θ = 4
(d-2) θ = 2

(a)

(b-1)

(b-2)

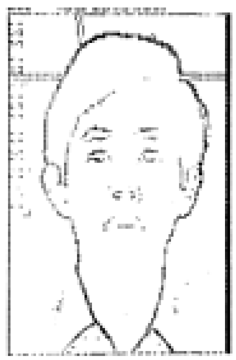
Example 2

(a) Gray-level picture

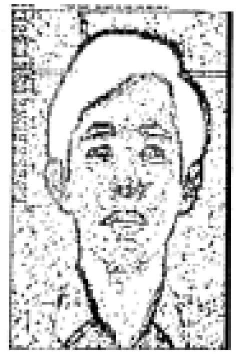
(b) Laplacian operator (Fig. 3-8)
 (b-1) θ = 25
 (b-2) θ = 35

(c) Robertz operator (Fig. 3-10(a))
 (c-1) θ = 4
 (c-2) θ = 2

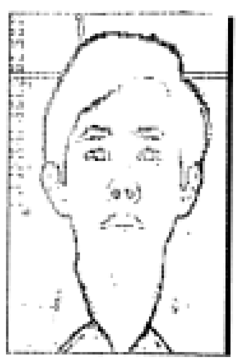
(d) Maximum of differences (Fig. 3-10(b))
 (d-1) θ = 4
 (d-2) θ = 2



(c-1)



(c-2)

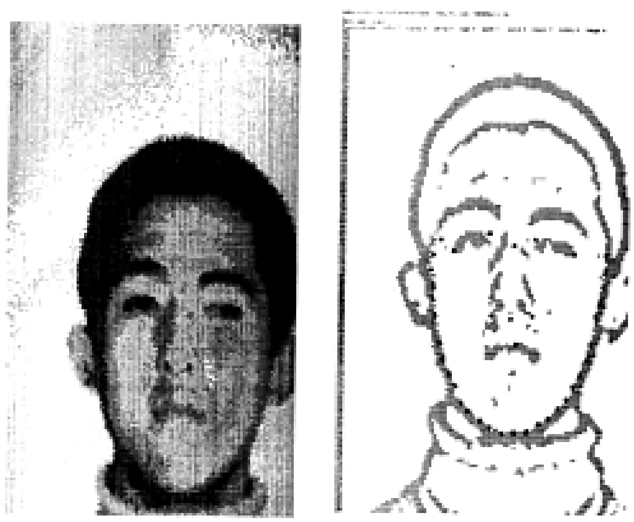


(d-1)

(d-2)

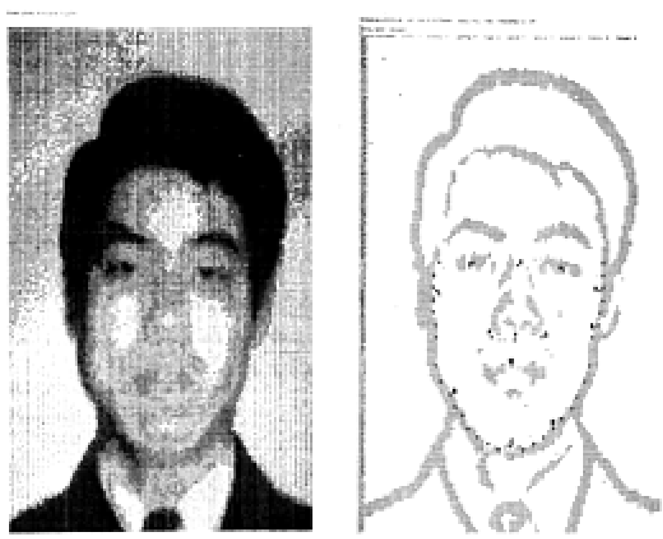
APPENDIX C

SUPPLEMENTARY DATA OF RESULTS OF ANALYSIS OF CHAPTER III

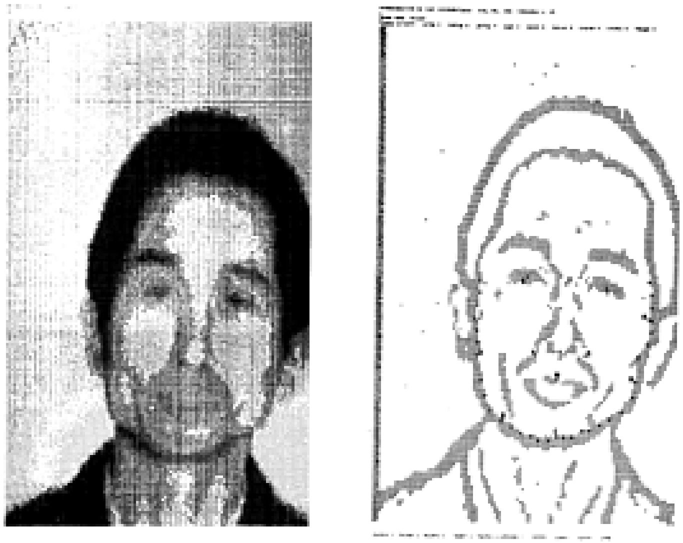


1. Simple Case

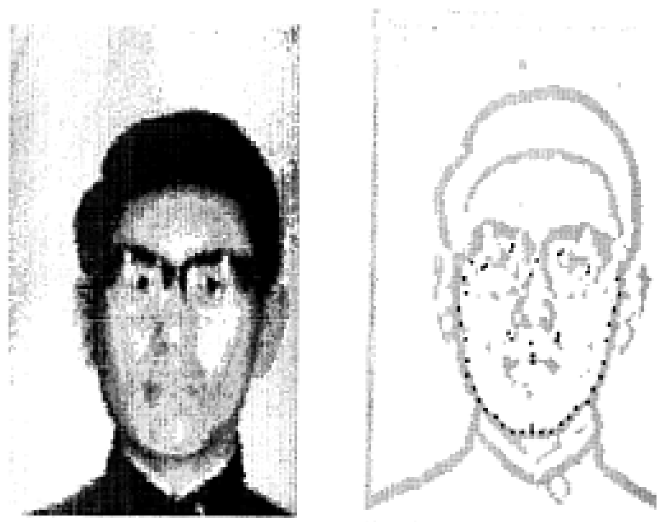
2. Chin contour with breaks



3. Separated eye components

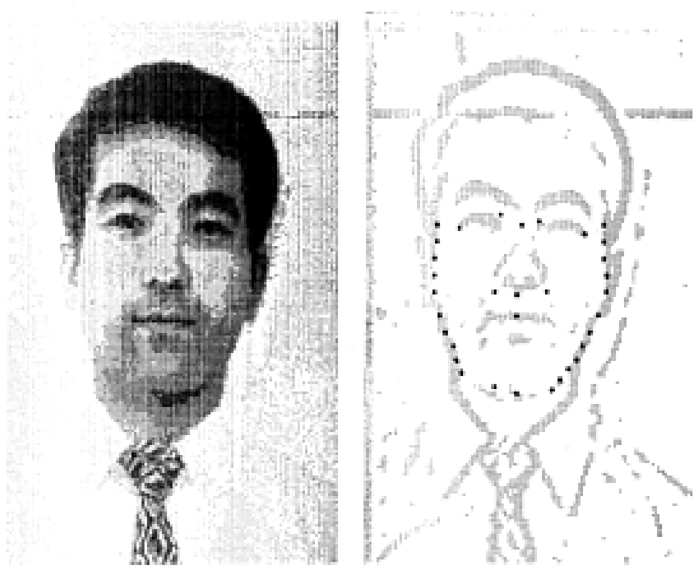


4. Face with facial lines



5. Face with glasses; eye centers happened to be obtained

6. Woman's face



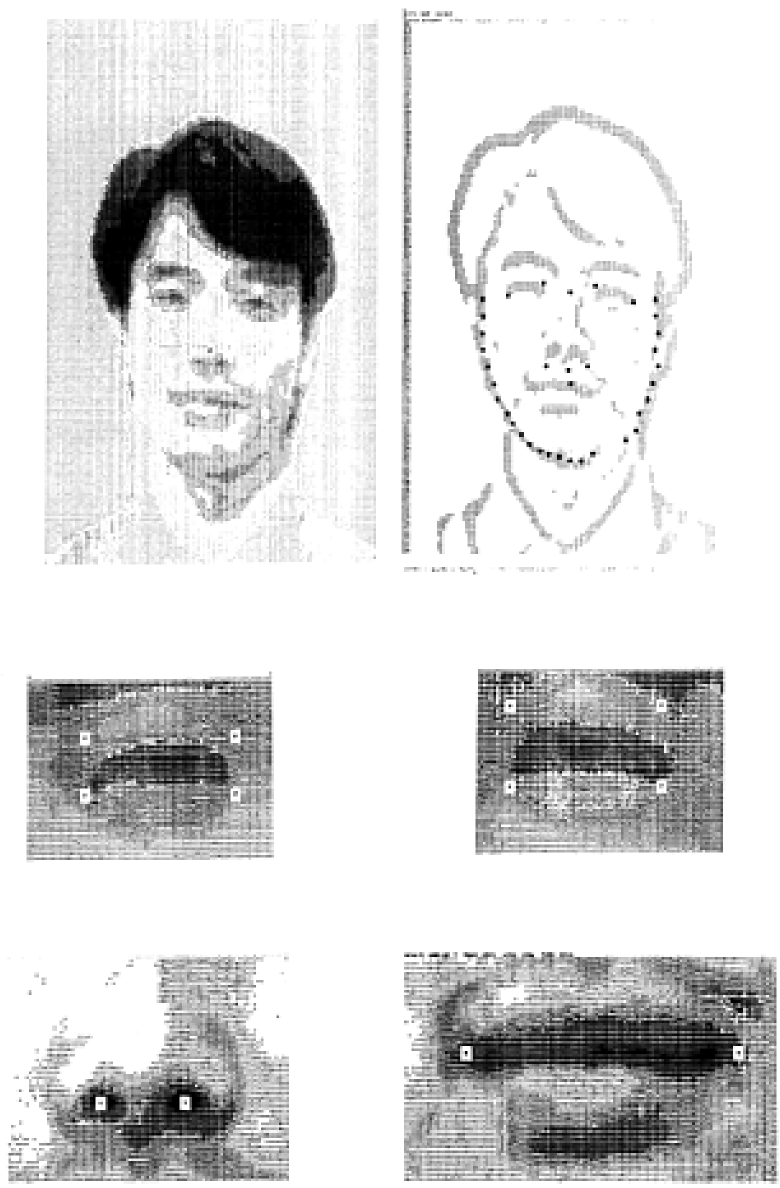
7. Sparse chin contour

APPENDIX D

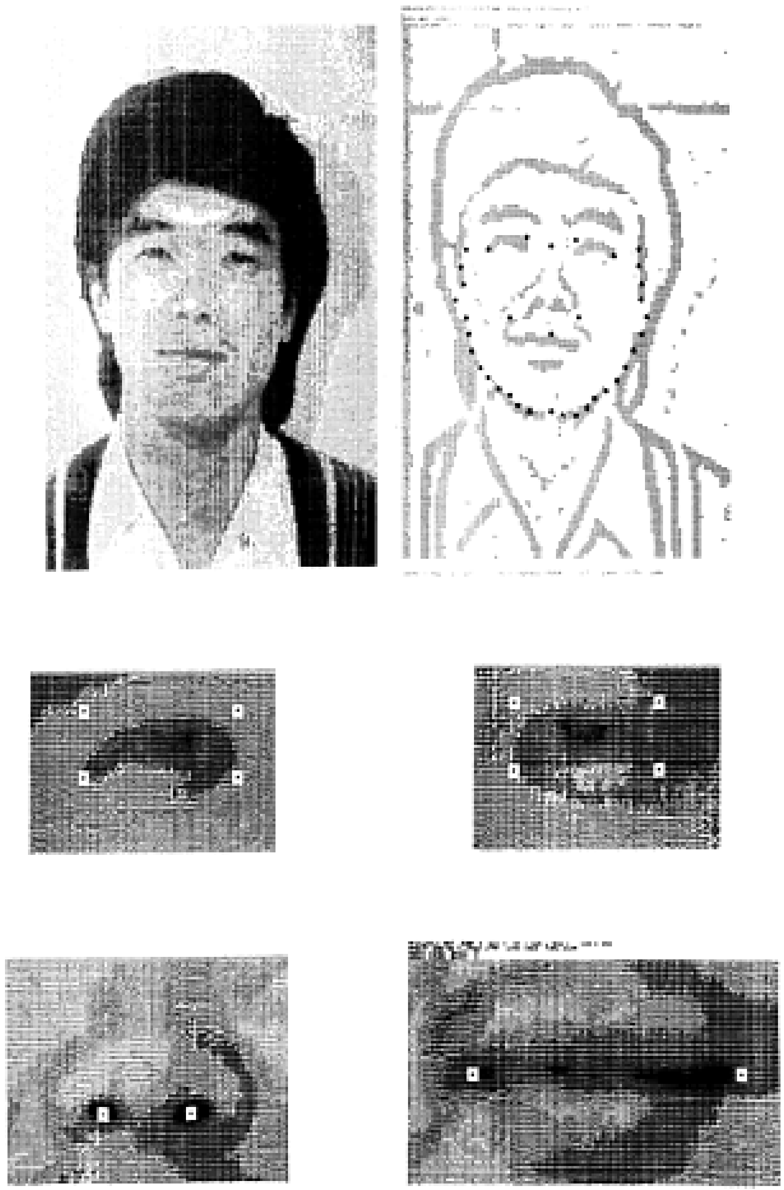
RESULTS OF FACE-FEATURE EXTRACTION OF CHAPTER IV

This appendix contains 10 sets of pictures (2 sets x 5 persons) to show the results of face-feature extraction described in CHAPTER IV. They are a part of 40 pictures used in the experiment of face identification. Each set comprises : (i) a digitized gray-level picture; (ii) a binary picture in which the feature points located in the first stage of analysis are marked as dots; and (iii) pictures displaying the right eye corners, left eye corners, nostrils and mouth extremities, which are located in the second stage of analysis.

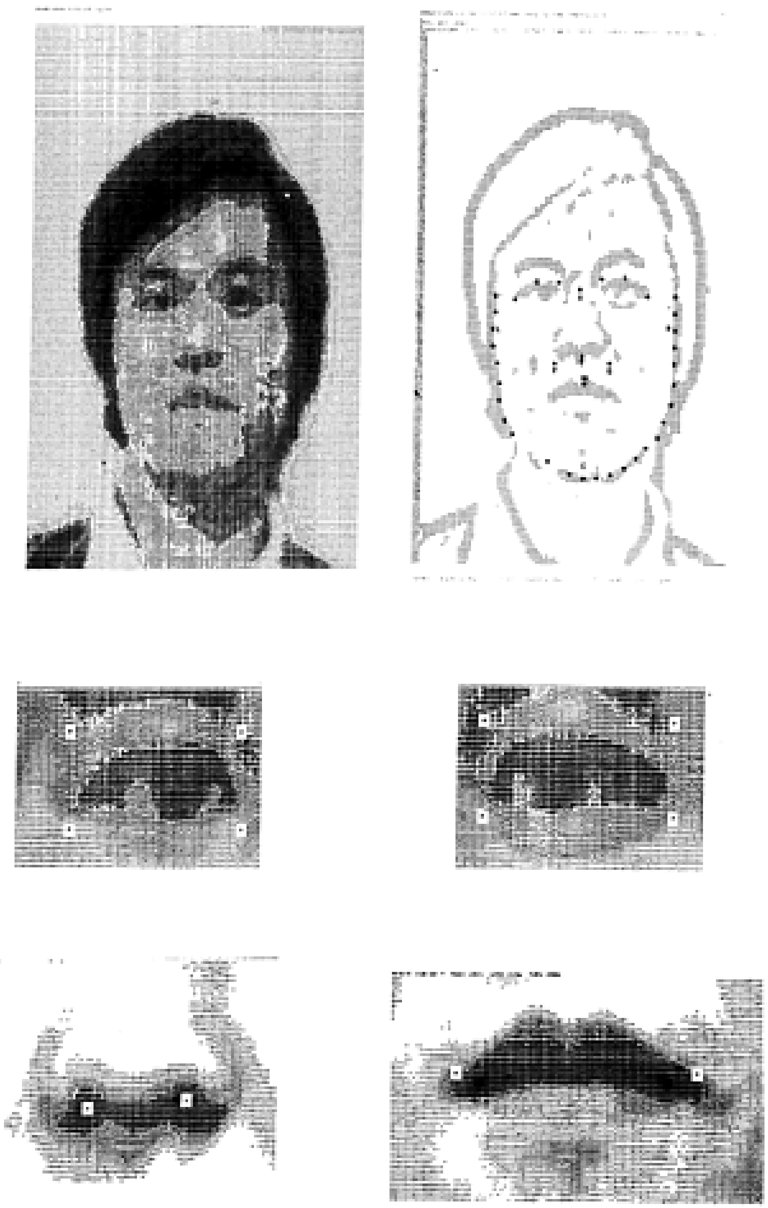
AMA1



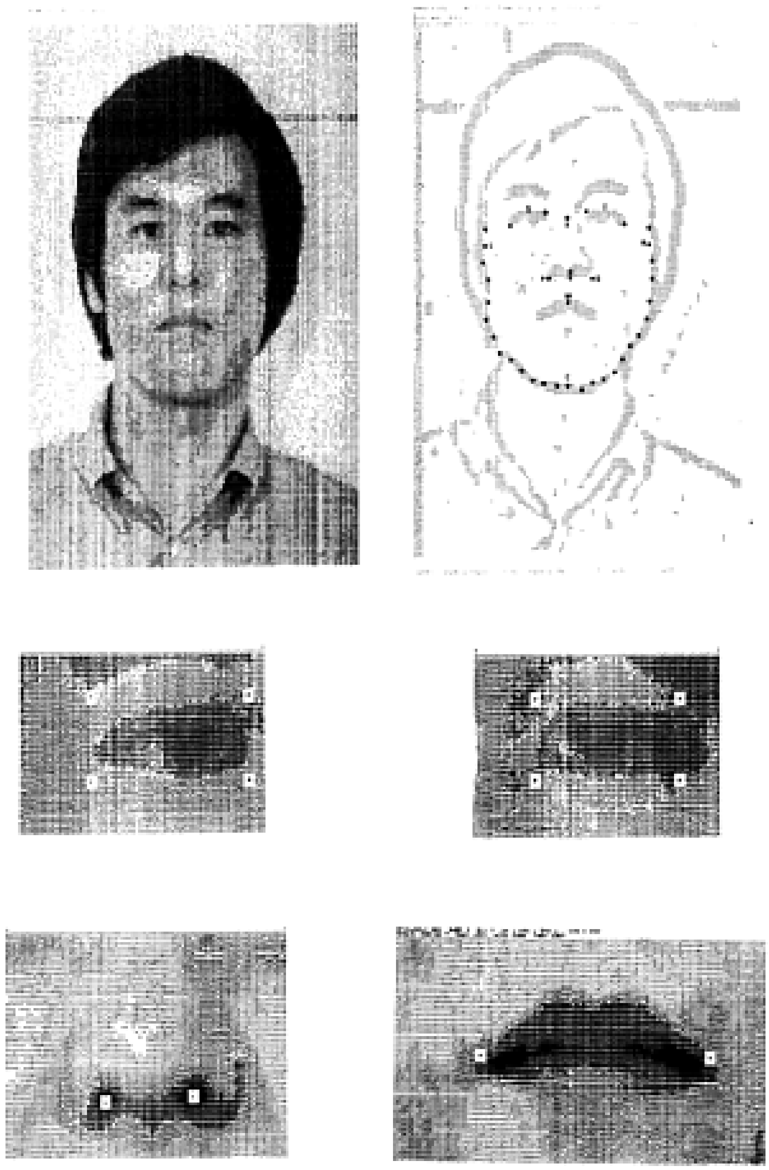
AMA2



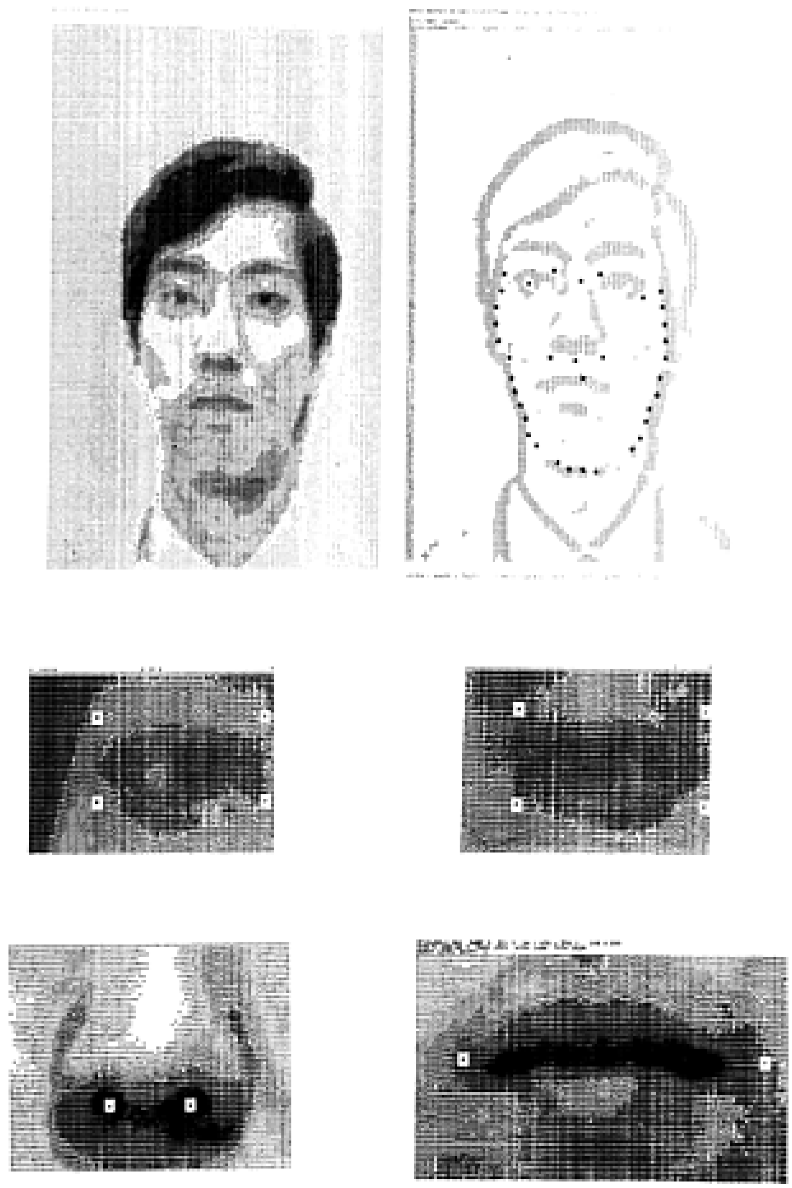
NAK1



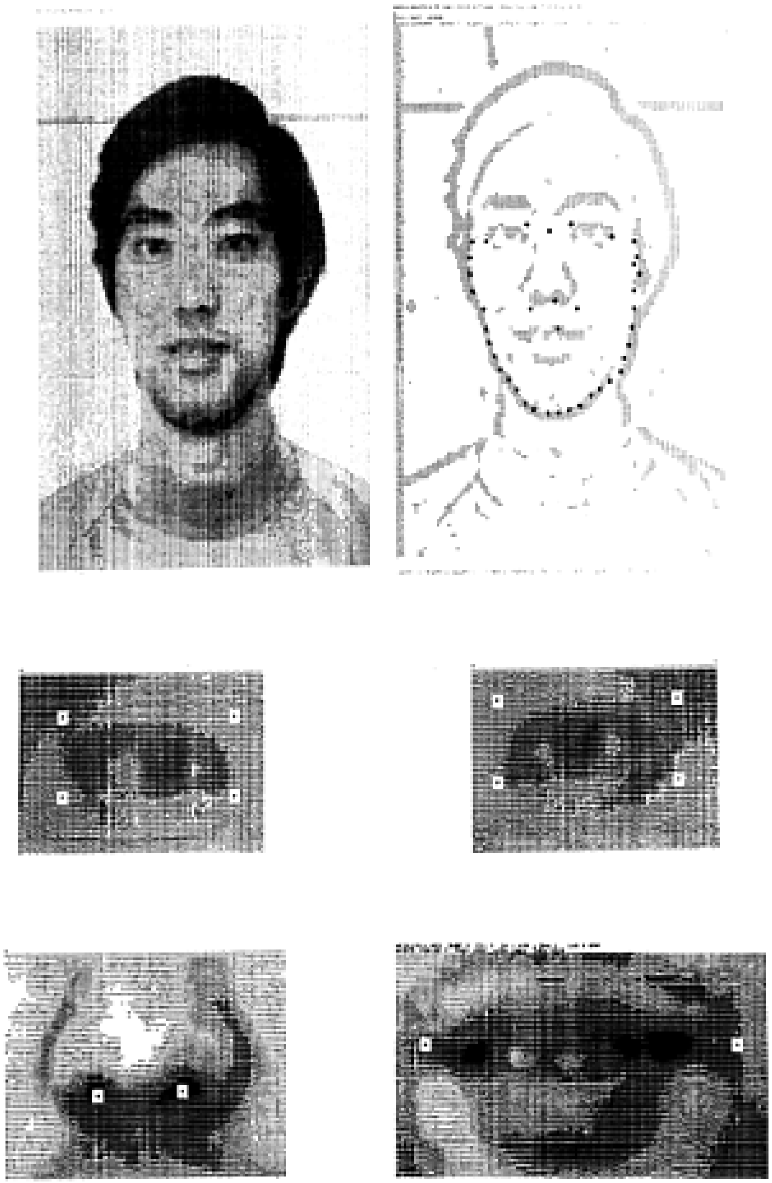
NAK2



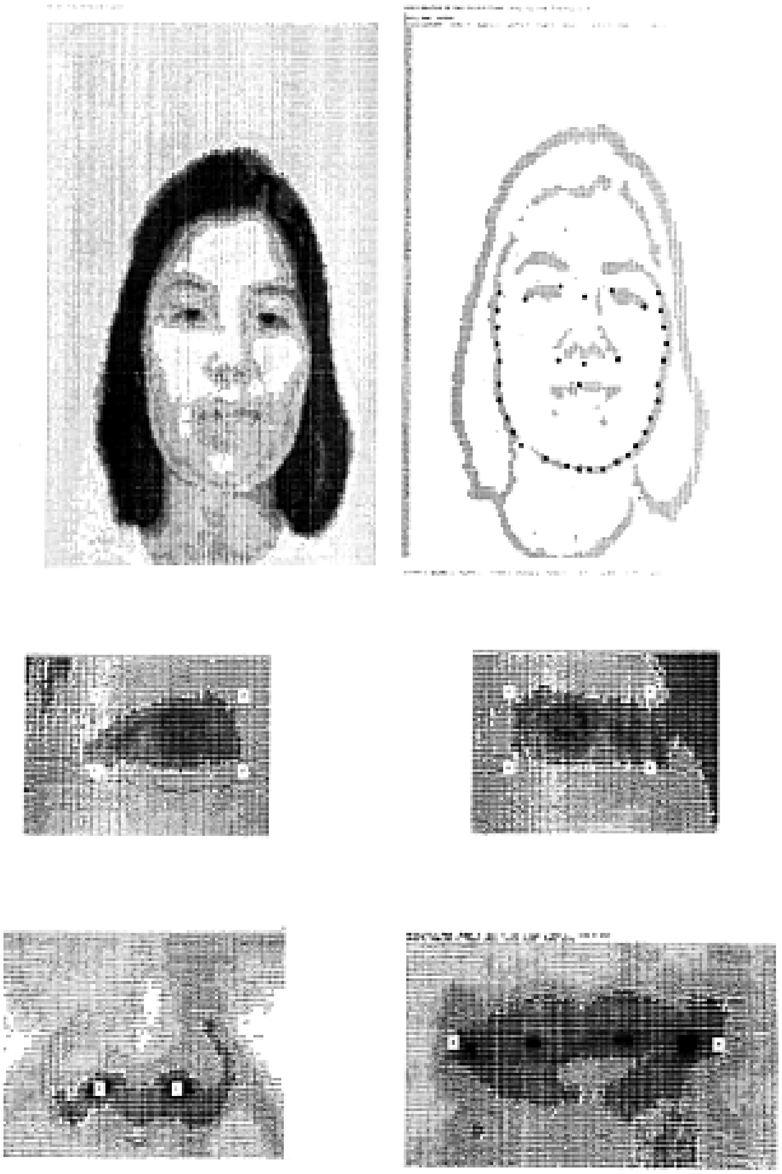
ON01



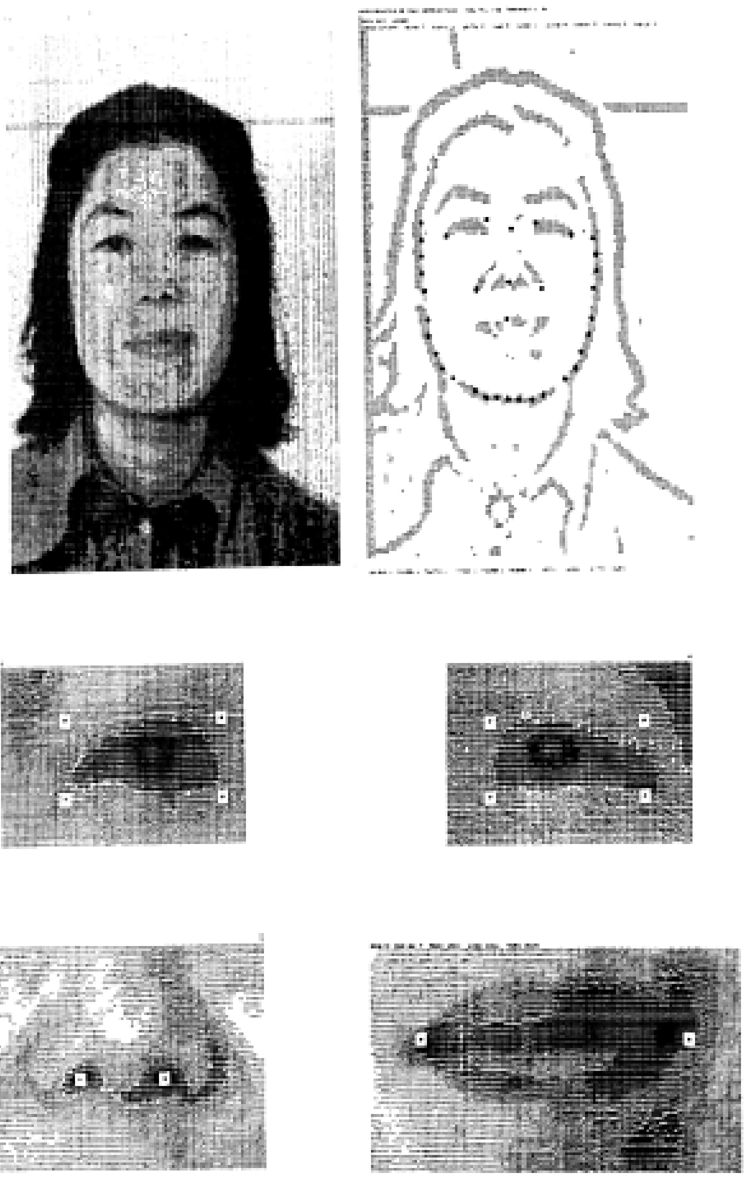
0N02



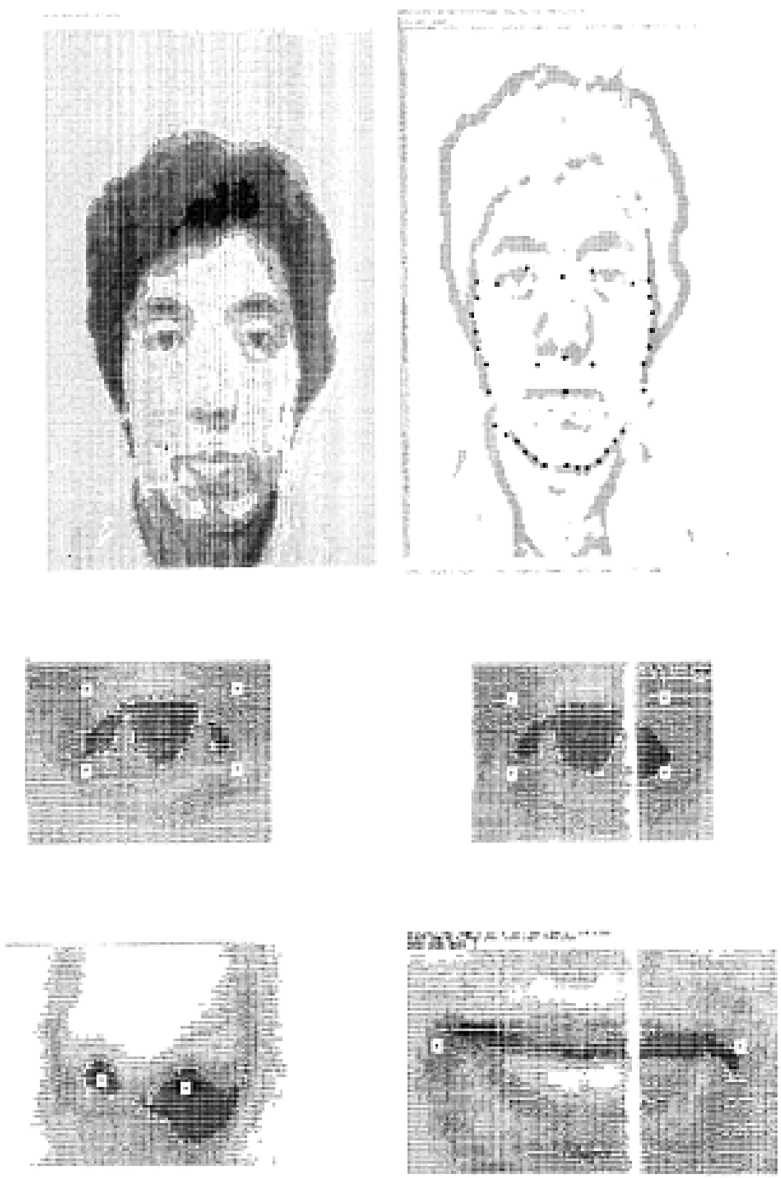
UEYT



UEY2



KONI



KON2