KAWASAKI STEEL TECHNICAL REPORT
No.27 (November 1992)
Hot-Rolled, Cold-Rolled and
Surface Coated Steel Sheets
and Electronics and Instrumentation

Image Processing System with a New Systolic Array LSI

Akira Ichinose, Kenji Suzuki, Yoshinori Wakimoto, Mitsuru Yanagisawa, Yuichiro Asano

---

Synopsis :
A LSI with a new type of systolic array was developed jointly by Kawasaki Steel and HNC, Inc. for image processing applications. This LSI is capable of performing convolution operations with a kernel larger than its own 8×8 array. The image size can be adjusted up to 4 096×4 096 pixels as required, which avoids the normal restrictions involved in processing images of various sizes and with variable-sized kernels. A new high-performance image processing system is now being developed to exploit the capability of this LSI in combination with a general-purpose RISC processor. Modular design for both the software and hardware gives the system outstanding expandability and flexibility of application.

The body can be viewed from the next page.

Image Processing System with a New Systolic Array LSI*

Akira Ichinose
Staff Assistant Manager, Engineering & Technology Dept., Integrated Systems & Electronics Div.

Kenji Suzuki
Staff Assistant Manager, Research & Development Dept., Integrated Systems & Electronics Div.

Yoshinori Wakimoto
Research & Development Dept., Integrated Systems & Electronics Div.

Mitsuru Yanagisawa
Manager, System Engineering Div. III, Integrated Systems & Electronics Dept., Kawasaki Steel Systems R&D Corp.

Yuichiro Asano
Staff Manager, Research & Development Dept., Integrated Systems & Electronics Div.

Synopsis:

An LSI with a new type of systolic array was developed jointly by Kawasaki Steel and HNC, Inc. for image processing applications. This LSI is capable of performing convolution operations with a kernel larger than its own 8×8 array. The image size can be adjusted up to 4 096×4 096 pixels as required, which avoids the normal restrictions involved in processing images of various sizes and with variable-sized kernels. A new high-performance image processing system is now being developed to exploit the capability of this LSI in combination with a general-purpose RISC processor. Modular design for both the software and hardware gives the system outstanding expandability and flexibility of application.

1 Introduction

In recent years, the image source used for image processing has become more diverse, and the demand has become stronger to be able to handle images of various sizes such as those from a scanner, high-definition TV (HDTV) signal, and microscope images, in addition to camera input with the conventional 512 × 480 pixels. It has thus become a major restriction when the image processing system limits the size of the image. In addition, the required speed for image processing is also strict, and it is considered desirable to complete the processing within a time proportionate to the area as the image size increases.

Two-dimensional convolution operations, which form the basis of image processing, are given by the following equation in respect of o_{mn}, the element for output of m-row and n-column:

$$o_{mn} = \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} w_{ij} * p_{m+i\,n+j}$$

The handled image is a two-dimensional array with i-row, j-column elements denoted by p_{ij}, and a two-dimensional weight matrix (kernel) whose i-row, j-column element is w_{ij} with a size of K-row by L-column.

Convolution with a large kernel has not been used very frequently, because there was no means to achieve a processing speed good enough for practical use, although such a demand exists. Conventional image-processing hardware and algorithms have mainly been produced for convolution operations with a kernel size of 3 × 3 (**Fig. 1**). Most of the hardware represented by Fig. 1 has shift registers for two lines and carries out computation after collecting the inputs necessary for convolution operations. Since this configuration uses shift registers, there is a restriction on the size of image that can be handled. Even if shift registers of the variable-length type are used, the image size is restricted by the maximum length of the shift register. Furthermore, if an attempt is made to expand for a larger kernel size, the scale of the hardware must be increased. Alternatively, when 3 × 3 hardware is repetitively used for large-scale-kernel operations, this involves a heavy load on the host computer, thereby making high-speed operation difficult. These restrictions are mainly caused by the methods used to access data.

In order to solve these problems and to offer a more

* Originally published in *Kawasaki Steel Giho*, 24(1992)1, 26-31

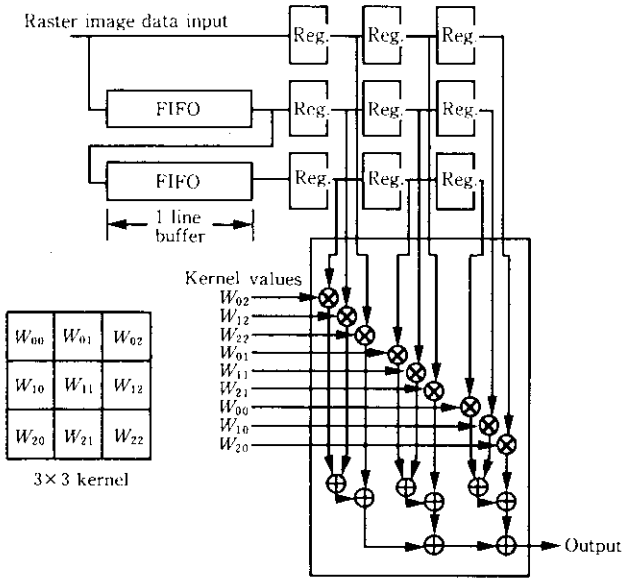

Fig. 1 Conventional 3 × 3 convolution hardware

flexible system, Kawasaki Steel has developed an image processing LSI jointly with HNC, Inc. of U.S.A. on the basis of architecture originally conceived by HNC, Inc. In addition, Kawasaki Steel is developing a new image processing system called "Dr. IMAGE II" to apply this LSI. Dr. IMAGE II reflects the following demands, which have been identified from the application of Kawasaki Steel's original "Dr. IMAGE," so that Dr. IMAGE II can be used for a broad range of inspection, measurement and recognition task that require image processing:

(1) Large images.
(2) High-speed processing of complicated calculations such as FFT and Hough transformation.
(3) Development and execution of image processing programs on the image processing system.
(4) Expandability of hardware (such as the addition of peripheral units).
(5) Easy connection to various types of work station.
(6) Stand-alone operation.

This paper describes the newly developed systolic array structure of the image processing LSI and then reports on Dr. IMAGE II which Kawasaki Steel is currently developing.

2 Systolic Array Structure of the Image Processing LSI

2.1 Background

Image processing is now being used in various fields, and camera input alone of a conventional 512 × 480 image is insufficient to handle all applications. An HDTV camera has 1 920 × 1 035 pixels, and microscopic images are formed from approximately 1 024 × 1 024 pixels. In the case of scanner input, it is necessary to handle even larger images. Instead of the conventional practice of determining the image size handled according to the image processing system available, the situation has changed so that the intrinsic demand is now to select the input source and resolution according to the purpose.

In addition, there are strict demands on the contents of processing. For instance, with neural network processing of images, which is now attracting considerable attention, convolution operations are frequently used. At this time, the size of the connection window between neuron layers, which corresponds to the kernel of convolution, is, in many cases, larger than the 3 × 3 kernel which has normally been used in image processing. There are two examples of using a large-scale convolution in the network model, one being Grossberg's "boundary contour system",[1] and the other being the "retina model for motion detection" by Eekman, Colvin and Axelrod.[2] The window sizes used in these examples are from 4 × 4 to 24 × 24, depending on the model and the layers of the visual system. As an extension of conventional image processing, it is also known that convolution of 7 × 7 or above is effective for structuring a process which is strong against noise. In this case, it is desirable to be able to process at high speed variable-sized large kernels.

2.2 Systolic Array

A systolic array is a mechanism arranged in units called "processing elements" which transmits data in a single direction and executes processing. The configuration of respective units is identical, and the contents of processing are the same. It does not matter whether there are many flows of data or whether the directions of data flow are different from one another. In addition, the arrangement can be made into a one-dimensional or two-dimensional array. Since the 1970s, numerous examples of systolic array architecture have been proposed and used for signal and image processing. In 1982, H.T. Kung described and classified many systolic arrays.[3]

An example of the 4 × 4 two-dimensional systolic array is shown in **Fig. 2.**

The systolic array architecture is very effective when multiple processors are used for handling processing with large calculation volume. For instance, multiplication of a matrix and filtering of one-dimensional/two-dimensional signals are suitable as direct applications of the systolic array.

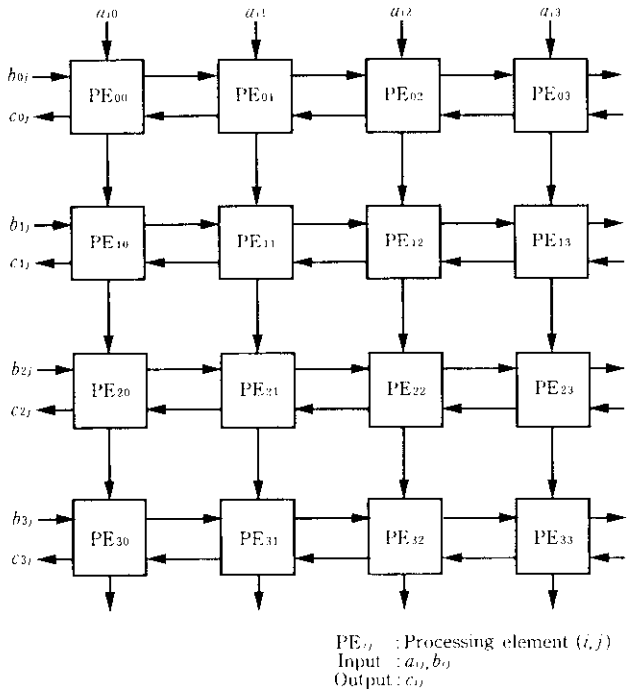

Fig. 2 An example of 4 × 4 systolic array

2.3 Systolic Array Structure for Image Processing

As an application of the new systolic array architecture proposed by HNC, Inc., the authors designed and produced an LSI device which can achieve convolution with a large kernel size using a single chip. On this single chip, and 8 × 8 array is integrated. When this LSI is used, processing an 8 × 8 convolution once requires one scan, and a 16 × 16 convolution requires four scans. Even when the kernel of the convolution and the window in the neural network are larger than the actual array, processing can be done without calculating or preserving an intermediate or partial sum. The major feature of this architecture is that the results at different positions can be directly calculated and output by each scan.

Therefore, it is possible to process the calculation of a large kernel by dividing it into several scans. For instance, when a 16 × 16 convolution is to be calculated with an 8 × 8 array, the output is calculated in the ratio of 1 to 4 for each scan, and this calculation is done four times by changing the position for the output.

In convolution with other kinds of systolic array architecture, it is necessary for calculation to provide the number of elements to match the kernel size, and connect these elements with the same structure as that of the kernel. The fact that hardware must be provided which corresponds to the size of the kernel to be actually processed is very inconvenient for use. The user of an image processing board employing such an LSI can obtain high performance, until the limit of the array size is reached. Once this limit is exceeded, however, performance suddenly drops. As a result, when the user develops an algorithm or makes application software, a smaller kernel is generally used to prevent the kernel from exceeding the size limit, thereby significantly increasing the processing time. For instance, it is known that when a kernel size of 7 × 7 or above is used in edge-enhancement algorithms, e.g. in the case of the Laplacian filter, noise can be effectively suppressed, although such a practice has not been used often, because it took too much calculation time. In general, an edge-enhancement algorithm using a large kernel suffers less from noise effects than that using a smaller kernel.

Further, the use of a neural network model of the human visual processing system has been limited to the execution software, or it was not used at all, simply because the hardware has been high in cost or there was insufficient ability to achieve a large window.

The present architecture eliminates these previous restrictions that have been imposed by the sizes of the kernel and window.

2.4 Outline of the LSI

The newly-developed LSI consists of a systolic array and a controller, which are used in combination. The systolic array LSI incorporates an 8 × 8 array, while the controller LSI incorporates an operator, look-up table (a value-conversion circuit by reference to the table), and address generator. Three memory blocks and one host bus can be connected for peripheral units. The LSI supports the function of directly transfering data to the host bus without intercession of the CPU. In addition, this data transfer and convolution can be executed in parallel. By being able to directly connect DRAM as an external memory, attention has been paid to realize a low-cost system. Each memory block can handle up to 16 MB maximum. The image size that can be handled is limited to 4 096 × 4 096 pixels only due to the restrictions imposed by the address generators and memory size. Image data are of the generally used raster type, and are stored in memory blocks, it also being possible to handle partial images. The size of each processed image is designated at every four pixels in the X direction and at every single pixel in the Y direction.

Furthermore, the size of the kernel for convolution has been selected so that from 8 × 8 to 64 × 64 can be handled by one chip set. When the kernel size is larger than 8 × 8, and high-speed processing is also required, multiple systolic array LSIs are used to configure an array having the actual size of the kernel being used, thereby making it possible to achieve the same speed as that for 8 × 8 with sizes up to 32 × 32 (16 LSI). Even in this case, it is sufficient to use only one controller LSI.

Table 1 LSI function list

One-dimensional convolution
Two-dimensional convolution
Matrix-vector multiplication
Gray scale transformation with look-up table
Constant operations—Add, Subtract, Multiply, Divide, AND, OR, XOR, NAND, NOR, XNOR (bitwise)
Inter-image operations—Add, Subtract, Multiply, Divide, AND, OR, XOR, NAND, NOR, XNOR (bitwise)
3 × 3 window binary image pattern matching
Comparison with constant
Count after comparison
Sub-image processing

Other significant functions of the newly-developed LSI are matrix-vector multiplication and one-dimensional convolution.

The basic functions of this LSI are summarized in **Table 1**, and its block diagram is shown in **Fig. 3**.

The processing procedure is next explained, using the execution of a two-dimensional convolution operation as the example. It is assumed that the image data to be used as the input for convolution processing is stored in a memory block controlled by the controller LSI. Namely, these data are those which have been transferred from the host bus or already stored as a result of previous processing. The processing proceeds as follows: (1) The image data are read from a memory block; (2) the values are converted by the input look-up table; (3) these converted values are sent to the 8 × 8 systolic array to carry out the convolution operation; (4) the output level is adjusted by the parallel shifter (level conversion circuit); (5) the output is converted by the look-up table; and (6) the result is finally written to another memory block. If the CPU side requires the processed results, data are sent from the memory block to the host bus.

2.5 Performance of the LSI

When all elements work as in the 8 × 8 convolution, this corresponds to 2 560 million multiplications and 2 560 million accumulations per second. By using the look-up tables in input and output pathes, binarization and gray-scale transformation of the images can be simultaneously executed. When the LSI is used for matrix-vector multiplication, only 16 elements operate out of the 64 (8 × 8) processing elements in the systolic array. This corresponds to 1 280 million arithmetic operations per second. Matrix-vector operation is only a supplementary function in image processing, but is important for neural network applications, particularly for a back-propagation network and other pattern-recognition purposes. In general, the basic processes carried out between two layers of the neuron are matrix-vector multiplication and succeeding non-linear function

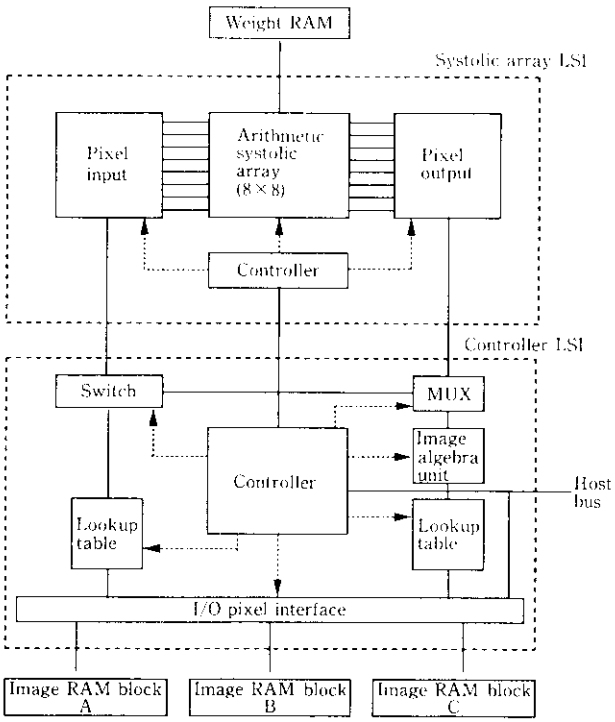

Fig. 3 Block diagram of systolic array LSI and controller LSI

Table 2 Performance comparisons—image processing time for convolution of 512 × 512 8-bit image (offered by HNC, Inc.)

(ms)

Kernel size	RISC work station (ms)	Company P LSI	Company I LSI	Company L LSI	HNC/KSC** Systolic array
3×3	2 000	6.6	13.1	13.1	6.6
8×8	14 000	26.2	6 chips	13.1	6.6
16×16	56 000	8 chips	18 chips	4 chips	26.2
32×32	224 000	*	60 chips	16 chips	104.9
64×64	896 000	*	220 chips	64 chips	419.6

* Not possible

** Not including DRAM access overhead

conversion. The use of the look-up table during output enables matrix-vector multiplication and non-linear function conversion to be simultaneously executed.

The LSI has very powerful image-processing ability, and **Table 2** shows a performance comparison of this LSI and the other LSIs on the market. As can be seen from the comparative table, the other devices have to be chosen so that the actual number of elements corresponds to the kernel size for the convolution. If an attempt is made to process a larger kernal size, processing becomes totally impossible or the processing time suddenly increases, because intercession and assistance from the host processor are required.

3 New Image Processing System, Dr. IMAGE II

3.1 Outline

Kawasaki Steel's general-purpose image processing system, Dr. IMAGE, was intended mainly research and inspection purposes, and incorporates the original image processing techniques which the company had developed. Experience of this system raised a number of additional demands for image processing: (1) the image size should not be restricted to that by conventional camera input and, in particular, larger images need to be handled; (2) complicated processing such as FFT needs to be executed at high speed; (3) the image processing system can operate by itself; (4) hardware can be expanded as and when necessary ; and (5) connection to various work stations should be possible. The necessity was also identified for a system which is flexible enough to expand the possibilities of image processing and recognition. By responding to these demands, Kawasaki Steel is developing a new image processing system, Dr. IMAGE II, on the basis of the new systolic array LSI. The standard system configuration is shown in **Fig. 4**.

Dr. IMAGE II, like the previous Dr. IMAGE, can be connected to a work station (WS). One of the important roles played by the current WS is to offer a highly-

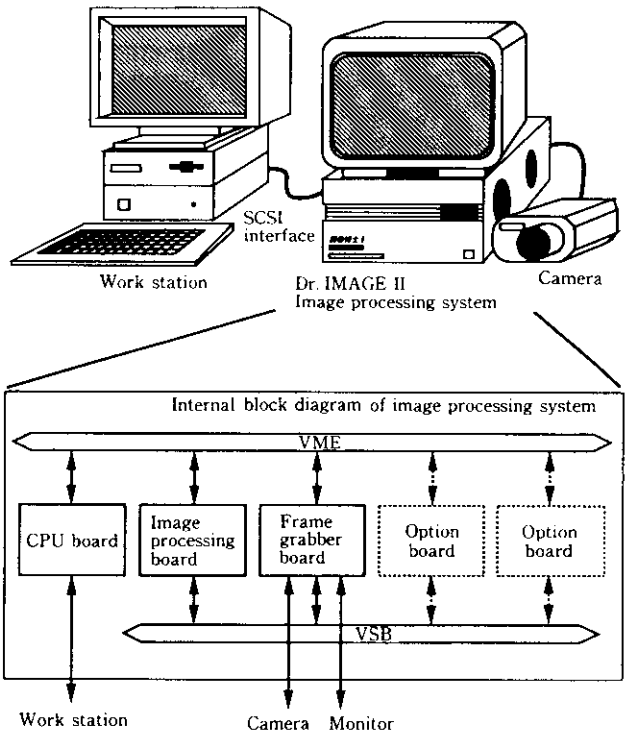

Fig. 4 Standard system configuration

developed human-machine interface. The present system uses a WS as the front-end for research and development of image processing and recognition. This arrangement enables the user to carry out operations on a familiar WS, and to structure a system without burdening the WS with calculation and without needing special hardware. The interface is a standard type, so that the system can be connected to many types of WS with little extra work. Furthermore, image processing and recognition is executed solely by Dr. IMAGE II; hence, the system can be operated independently and does not depend on the WS.

When conducting image processing independently of

a WS, control is transferred to Dr. IMAGE II. The LSI can be considered as a unit for special purposes in the system.

To permit system expansion, an internal system bus for standard use is provided, together with vacant slots. Consequently, peripheral units for functions closely related to image processing can be connected to or incorporated into the system by using these slots.

Image input/output is handled by a standard 512 × 480-pixel 24-bit full-color board. Since this board is a module that connects to a standard system bus, the board can be changed to suit larger images or special image input/output.

The features of the newly-developed LSI are actively exploited, making it possible to designate, within the range from 8 × 8 to 2 048 × 2 048 pixels, every four pixels in the X-direction and every single pixel in the Y-direction (the maximum image size that can be processed by the LSI is 4 096 × 4 096 pixels). A 6.5-ms processing time has been realized for the 8 × 8 convolution of a monochrome image with a standard input size of 512 × 480 pixels.

Complicated processing such as fast Fourier transformations (FFT) can be done without burdening the host computer, because a reduced instruction set computer (RISC) is provided for exclusive use. In addition, the user can independently develop special software.

When operating through the WS, three interfaces of "menu," "shell" and "library" are provided. To operate Dr. IMAGE II, the menu is displayed graphically in a window and operated with a mouse, the shell uses commands from the keyboard, and the library is called from the user program.

3.2 Hardware

The WS and Dr. IMAGE II hardware are connected by a small computer system interface (SCSI) bus. In Dr. IMAGE II, a VME system bus connected to the CPU board and image processing board forms the basic architecture. Beside these boards, a frame grabber board is normally connected. Dr. IMAGE II has, apart from the system bus, a VSB image bus for exclusive use in image transmission. This VSB bus connects the image processing board and the frame grabber board.

The CPU board controls all the operations of Dr. IMAGE II. This board is also provided with a SCSI interface, which is a general-purpose bus, and is connected to the WS to hand over commands and data.

Image processing board takes charge of most data processing for images, and is provided with the new image processing LSI, the RISC processor operating at 40 MHz and a memory. The RISC processor manages and controls all the components on the board, including the image processing LSI, and also processes data. Since this is a general purpose processor, a user can develop programs that work on this board.

The frame grabber board inputs the images from a camera and displays them or processed images on a monitor. To ensure a continuous input and to prevent any disturbance of the display, input and output are each provided with double data storage memories.

Table 3 Hardware component function list

Work station	Human-machine interface Network interface Data storage
CPU board	Image processor control General-purpose I/O interface System resource management
Image processing board	Signal (image)/data processing Image bus control
Frame grabber board	Video signal digitizing Image data temporary storage Image display

The system can be expanded by connecting a new board to the system bus. In the case of a board for handling images, an image bus other than the system bus may be provided. System expansion is also possible by connecting a peripheral unit to the system through the CPU board or through an additional interface board.

The functional assignments are summarized in **Table 3**.

3.3 Software

The standard software architecture, when the system is connected to a WS, is shown in **Fig. 5**. The WS side and Dr. IMAGE II side are basically symmetrical. By making the management methods for two systems iden-

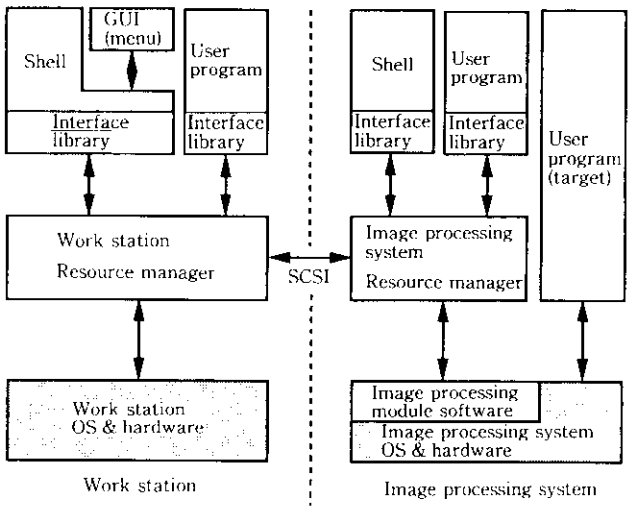

Fig. 5 Software architecture

tical, the architecture could be simplified and independent operation of Dr. IMAGE II became possible. Communication between the WS and Dr. IMAGE II are carried out between two management programs. The communication procedure consists of transmitting a command to one system as a request for processing to utilize resources not registered in the other system. This makes it possible to handle changes in the system elements. In addition, the same command is used in communicating between other software elements, thereby facilitating future expansion. The library which generates the command is supported, thereby permitting direct processing by the program.

Interactive operation with the user is carried out by the "shell" software. When the system is being operated with a window through the mouse, the Dr. IMAGE II menu, namely the graphical user interface (GUI) for image processing, is operated on the X-window system of the WS. The menu generates commands to the shell, and utilizes the services of the shell to simplify the system. In order for the user to use Dr. IMAGE II from a program written in C language, an exclusive-use interface library of Dr. IMAGE II is used.

The shell or the user program employs the interface library, thereby communicating to the system management software, called the "resource manager," by using the same communicating format to instruct in the execution of a command. The resource manager searches for the required resources from those that are usable under its control such as the image processing board and I/O unit and executes the command. The actual handing-over of the command is also executed on this board, since the image processing software is operating through image processing board. The purpose of the image processing software is to execute a program that has been developed by the user on this board at high speed. As an environment for developing a program by the user, a C compiler and debugger/simulator will be provided.

In order to operate Dr. IMAGE II independently, communication with the WS is cut off, and operation is carried out by using the shell in the system or a user program. The system can be operated without the help of the resource manager from the user program during independent operation.

4 Conclusions

The hardware restrictions, which a conventional image processing system has imposed on the image size, kernel size for convolution and processing time, have been removed by a new image processing LSI. This device is being incorporated into a new image processing system, Dr. IMAGE II, which has the following features:

(1) The image size and kernel size are variable so that a large image size (up to 4 096 × 4 096 pixels for the LSI and up to 2 048 × 2 048 pixels for the complete system) and a large-scale kernel (up to 64 × 64) can be handled.
(2) The 8 × 8 convolution of a monochrome image of 512 × 480 pixels can be executed in 6.5 ms, the processing time being proportional to the image size and convolution size.
(3) In addition to the CPU for controlling the system, a RISC processor is incorporated, and the system can be operated independently or connected to a WS, using software that can handle changes in the system elements.

The development of the new LSI device has enabled convolution operations with a large image size and large scale kernel to be conducted, which have previously been impractical due to restrictions on processing time. It is expected that new algorithms and application programs for use in these processing operations will be developed in the future. Furthermore, the LSI is used in combination with a general-purpose RISC processor, so that complicated processing is possible without burdening a host computer, thereby utilizing the system hardware to its maximum.

Dr. IMAGE II is a system with outstanding expandability and flexibility in its application, and offers a high operating speed as a result of incorporating both software and hardware modules and its clear-cut functional assignment. Dr. IMAGE II can be widely used in developing the basic algorithms needed for image processing and in the fields of application studies, inspection, measurement and neural network processing of images.

5 Acknowledgement

The authors express their deep gratitude to HNC, Inc. and to Dr. Robert W. Means who conceived this architecture, and who has given his kind and valuable cooperation in supplying reference materials and compiling manuscript for this report.

References

1) S. Grossberg and E. Mingolla: "Neural Dynamics of Perceptual Grouping: Textures, Boundaries and Emergent Segmentations", *Perception and Psychophysics*, **38**(1985), 141-171
2) F. H. Eekman, M. E. Colvin, and T. S. Axelrod: "A Retina-Like Model for Motion Detection", International Joint Conference on Neural Networks, Washington, D.C. (USA), (1989), II-247-249
3) H. T. Kung: "Why Systolic Architectures?", *Computer*, 15(1982)1, 37-46