MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

Advanced Vision Processor with an Overall Image Processing Unit and Multiple Local Image Processing Modules

Hiroaki Kubota, Kazuhiro Fukui, Minoru Ishikawa,
Hiroshi Mizoguchi, and Yoshinori Kuno

Research and Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210, JAPAN

ABSTRACT

This paper proposes a new concept for image recognition and understanding whereby both overall and local image processing are executed in parallel. Using this concept, we have developed an advanced vision processor that can recognize moving objects and handle moving pictures.

The vision processor carries out feature extraction for image recognition by means of several local processors in parallel from time varying images on which an overall image processing unit has performed low-level operations such as data conversion and filtering. Special characteristics of the system are that each local processor processes only its region of interest, the division of regions being determined from the image feature and contents.

A balloon-juggling robot has been developed to demonstrate the capability of the system. The system recognizes the balloon and measures its 3-D position in real time. The robot hits the balloon by using the 3-D data.

1. Introduction

Image recognition and understanding are divided into three levels[1]. The first level is low-level processing that transforms image data into image data. The second level is intermediate-level processing that converts image data into symbolic data. The final level is high-level processing that manipulates symbolic data to obtain recognition results. Various operations are necessary at each level. Achieving of high-speed processing requires that all operations be processed at high speed. The overall speed of processing will be low, if one part of the processing is slow even though the rest is fast.

Many image processing systems have already been developed, and many systems are commercially available. Most of these systems have special hardware for low-level processing. For example, TOSPIX[2] uses a local parallel processing module and WARP[3] has a systolic array that executes pipeline processing. These systems are useful for applications in which low-level processing plays the main role and only simple processes are needed at intermediate and high levels. However, they are slow for applications which need sophisticated intermediate and high level processes to recognize complex scenes, because such processes run sequentially on the host computer.

We aimed at developing a system that understands moving images. The system has a special hardware architecture that is suitable for all three levels of image recognition. The Image Understanding Architecture (IUA) [4] is also a system with three level processors. Though this is an ambitious project, the scale of the system is too large to be used for robot vision.

In designing a vision system, we must carefully consider the image input issue. The overall performance might be limited even with fast computing if the overhead for image input is large. The Multiple Window System[5] is a solution to this issue. It has several image buses through which multiple processors can obtain any region of the time-varing images. However, the three-level hierarchy of image understanding is not considered. Each processor in the system covers all processing levels. Thus, good performance cannot be expected if an application needs time-consuming low-level processing, such as spatial filtering.

We propose a vision system taking into consideration the three levels of image understanding and the image input issue. The system consists of an overall image-processing unit and multiple local image-processing modules, all connected with special image buses. The overall processor contains special hardware for low-level vision using local parallel processing and pipeline processing. It carries out low-level processing in real time. The multiple local processors are assigned intermediate-level. A variety of parallelisms inherent in image processing are used in dividing and assigning tasks to each local processor to speed up intermediate-level vision. High-level processing is executed by a host workstation since the amount of symbolic data at a high level is small.

This paper presents the system configuration and experimental results of moving image recognition as demonstrated by a balloon-juggling robot.

2. Architecture

2.1 System Configuration

Fig.1 shows the system configuration of the Vision Processor (VP). The following two issues were considered in system design:

- Easy image input and handling of moving pictures
- Suitable processing for the three levels of vision

The system consists of an overall image-processing unit and multiple local image-processing modules. The overall processing unit consists of special hardware for local parallel processing and pipeline processing and can perform low-level vision at video speed. Each local processing module incorporates a micro-processor and performs several operations at the intermediate-level vision with software. The host computer is an engineering workstation (EWS) and this handles high-level vision. The system has several special image buses and can transfer image data between the overall processing unit and local processing modules in real time.

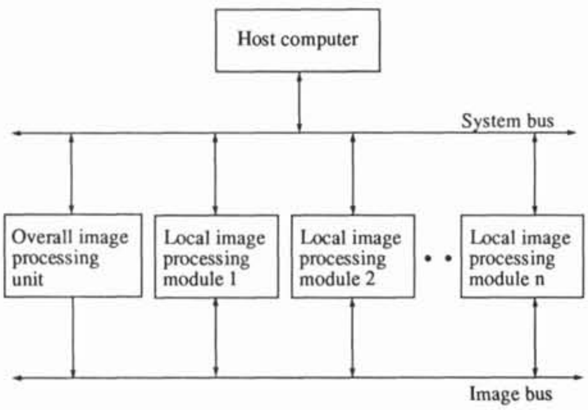

Fig.1 System configuration of vision processor

2.2 Parallel Processing in Local Modules

The conventional way of assigning an image region to each local processor in a multi-processor system is to divide an image with a regular interval grid. In order to carry out intermediate-level vision more efficiently, several ways have been devised to use the local modules in the new system.

In parallel processing by multiple processors, data communication between processors would ordinarily slow down the processing speed. This system assigns each local module depending on the image content to avoid this issue. Each module attends to a region containing some interesting features detected by the overall image-processing unit or other local modules. Since most processing in each processor is confined to one region, the amount of communication between processors is reduced.

One of the methods of division depending on the image content is to assign each local module to a circumscribed rectangular region of a labeled connected component in a binarized image. Fig.2 shows an example of this type of division. The method is useful for feature extraction operations in each connected component.

Another interesting method is to determine the search region by using object models or features extracted from an image. Such information is used to select candidate regions, to each of which a local module is assigned. In Fig.3, the circular regions are extracted by a simple matching method running on the overall image-processing unit. Each local processing module attends to the rectangular region enclosing one circle.

In image processing, top-down analysis is often adopted using object models and other a priori knowledge of the objects. Several candidate regions are hypothesized where target features should exist. In this case, each local module searches the selected regions for the target features.

The local image processing modules can be used in a few other ways. One is in a multiple resolution or

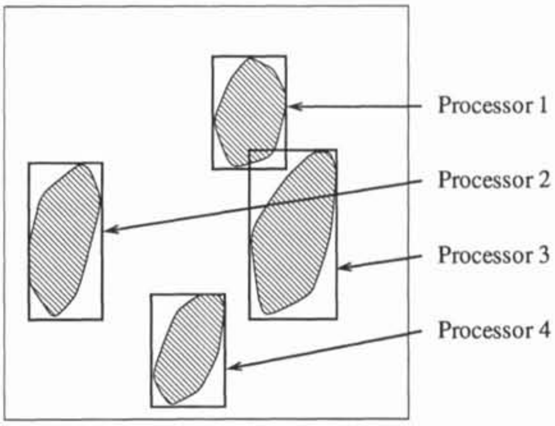

Fig.2 Region division depending on image components

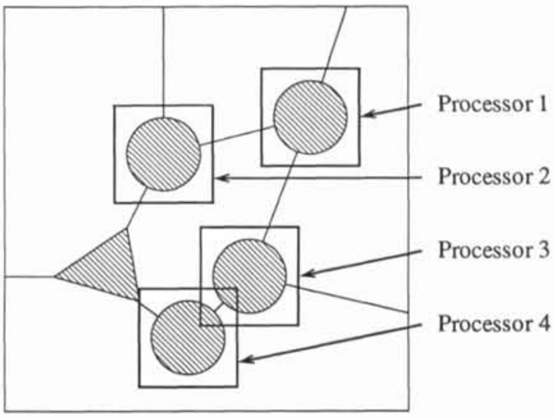

Fig.3 Region division depending on image feature

an image pyramid way. A local module can receive part of an image through image buses at any resolution, such as full resolution or half resolution taking every other pixel. An image can then be analyzed by several local modules, each at a different resolution. The local image processing modules can also be used in the manner of Multiple Instruction Single Data (MISD). Several local modules apply different operations to the same region of an image.

The new system realizes these parallelisms on several local modules by instructions from the host computer.

3. Vision System with 16 Local Modules

We have developed a vision processor using the concept described in Section 2. The system consists of an overall image-processing unit, sixteen local processing modules, an image input module, an image output module, six special image buses, and a host computer, as shown in Fig.4. An engineering workstation AS-3160M (TOSHIBA) is used as the host computer. It also serves as the machine for program development. A general purpose image processor TOSPIX-i (TOSHIBA) forms the overall image processing unit. It processes all pixels of a given image at video rate and chooses the interesting regions of the image.

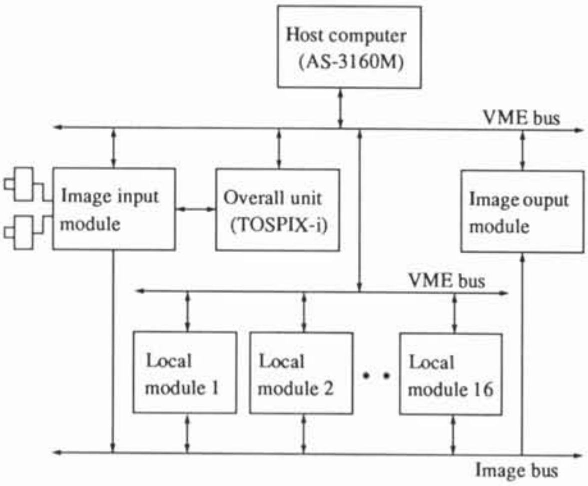

Fig.4 Developed vision system

Each local module consists of a micro-processor M68030, a window controller, three planes of window memory, an address decoder, and a bus interface, as shown in Fig.5. The module obtains image data for a specified interesting region from the image buses within 16 ms. The window memory which holds the interesting region has three planes for moving image processing and color processing. The image buses also consist of six sets: three for image input and three for image output to allow color image processing. Parallel data ports are included in the bus interface in

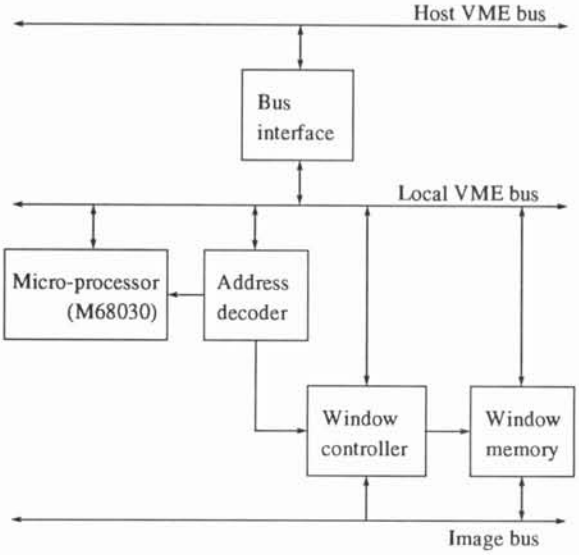

Fig.5 Local image processing module

order to receive information on the interesting regions and the results of intermediate-level processing without stopping of local processing.

The image input module provides two video input ports for stereo vision. The module also has a function for specified color extraction. The image output module can show the results of all local modules on a display monitor.

4. Experimental Results

A preliminary experiment was made to check the capability of the system. In the experiment, the overall processing unit extracts several interesting regions from a given image and the local processor modules calculates the center of gravity of the objects in the interesting regions. The system transfers the input image data to the overall processing unit. The unit filters the data, and carries out binarization, region labeling, calculation of the circumscribed rectangular regions, and pixel counts for each connected component in order, and determines the interesting regions. By using pipeline processing, the

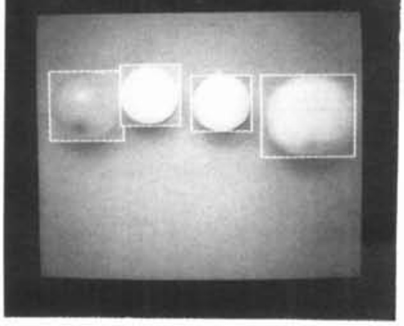

Fig.6 Result of determining interesting regions

unit carries out this procedure in 33 ms. The host computer transfers position information on the interesting regions from the overall processor to each local processor and distributes the image data to the local processors through one of the image buses. The chosen interesting regions are shown in Fig.6.

On the other hand, when information on an interesting region is received, each local module places the image data into the window memory and carries out feature extraction operations. In this experiment, each module checks the size of the connected component and calculates the coordinates of its center of gravity. Each module sends the result to the host computer through the parallel ports. The processing time on each local module in this simple experiment was about 20 ms, depending on the size of the interesting region. The experimental results have proven the capability of the system in terms of its basic image processing function.

To demonstrate moving picture analysis capability of the system, the vision proceesor was combined with a seven-degrees-of-freedom direct-drive arm[6] to create a balloon-juggling robot. Using two TV cameras, the robot ascertains a balloon's movement three-dimensionally at a sampling rate of 10 frames per second. It then determines the arm movements necessary to put the racket in the estimated striking position within the calculated time. Fig.7 shows a photograph of the robot.

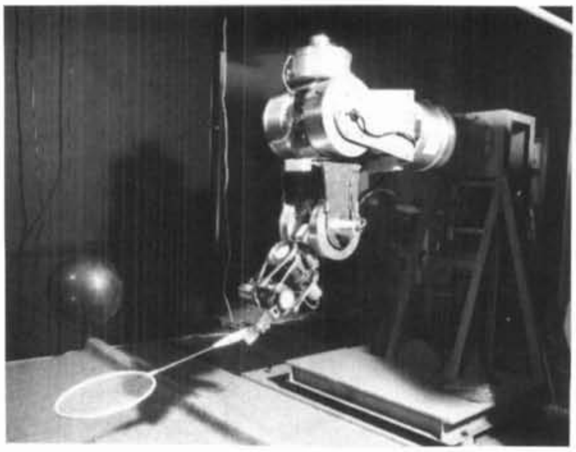

Fig.7 Balloon juggling robot system

In this robot system, the overall image-processing unit selects the color regions for the balloons in the scene from the moving images. Each local module is ordered to examine certain color regions of the left or right images such that a particular local module attends only the red regions of the left images to track a red balloon in the left camera. Thus, the vision processor can recognize several different color balloons at the same time, though the robot could not juggle two or more balloons in this experimental system. In order to juggle many balloons, a sophisticated robot control program will need to be developed in addition to the recognition capability realized in this experiment.

The experiment was successful; the robot was able to keep hitting a balloon more than a hundred times. This result has proven the moving image analysis capability of the system.

5. Conclusion

This paper has proposed an advanced vision processor for moving image processing. The concept of this vision processor is that overall and local image processing are executed in parallel and that local processors take up various parallel computation styles according to the processing procedure for intermediate-level vision.

A balloon-juggling robot was developed based on the vision processor. The result, where the robot continued to hit a balloon more than a hundred times, demonstrated the moving image processing capability of the system.

The separation of overall and local image processing realizes higher processing speeds and enhanced flexibility. With different software on the local modules, this vision system is expected to find many applications, such as in factory automation systems and robot control systems.

REFERENCES

[1]Tanimoto, S. L., "Architectural issues for intermediate-level vision," ***Intermediate-Level Image Processing***, Duff, M. J. B., ed., Academic Press, London, 1986.

[2]Kidode M. and Shiraogawa Y., "High-speed image processor : TOSPIX-II," ***Evaluation of Multi-computers for Image Processing***, Uhr, L., et al., ed., Academic Press, 1986.

[3]Gross, T., Kung, H. T., Lam, M., and Webb, J., "Warp as a machine for low-level vision," **Proc. of 1985 IEEE International Conference on Robotics and Automation**, pp.790-800, 1985.

[4]Weems, C. C., Levitan, S. P., Hanson, A. R., Riseman, E. M., Shu, D. B., and Nash, J. G., "The Image Understanding Architecture," ***International Journal of Computer Vision***, Kanade T. and Brooks,R., ed., Vol.**2**, No.**3**, pp.251-282, Kluwer Academic Publishers, 1989.

[5]Inoue H. and Mizoguchi H., "A flexible multi window vision system for robots," **Proc. of 2nd International Symposium of Robotics Research**, pp.42-49, 1984.

[6]Hashimoto H., Ozaki F., Asano K., and Osuka K., "Development of a pingpong robot system using 7 degrees of freedom direct drive arm," **IEEE Proc. of IECON '87**, pp.608-615, 1987.