Taylor & Francis
Taylor & Francis Group

Advanced Robotics

ISSN: 0169-1864 (Print) 1568-5535 (Online) Journal homepage: https://www.tandfonline.com/loi/tadr20

Vision-based mobile robots on highways

Ichiro Masaki

Advanced Robotics, Vol. 9, No. 4, pp. 417–427 (1995)

Vision-based mobile robots on highways

ICHIRO MASAKI

Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Room NE43-721, 545 Technology Square, Cambridge, MA 02139, USA

Received for *AR* 31 May 1994

Abstract—Intelligent vehicles are mobile robots on highways. They are expected to improve the safety, efficiency and environmental impacts of the current highway traffic systems. Vision systems will play an important role as sensors for the intelligent vehicles. This paper first compares the vision sensors with other sensing methods from an application point of view and then describes two vision systems, one which we have developed and another which we are developing. Two important features are required for the vision systems applied to intelligent vehicles: three-dimensional (3D) measurement capability and real-time operation. We chose a trinocular stereo vision scheme among a number of 3D vision processing methods because it is suitable for real-time operations with dedicated processor architectures. The trinocular stereo algorithm requires a large number of operations, but all the operations are relatively straightforward and, therefore, they are suitable for custom architecture implementation. The system takes three images simultaneously by using three TV cameras installed on a single horizontal line at the front grill of the test car. Vertical edges are extracted from these images and the spatial offsets (or disparities) among the images are calculated for measuring the distances to the objects. The first version was developed and installed in a car for highway testing. Two custom digital processors were developed: one for edge detection and the other for stereo matching. The test results were encouraging and the architectures based on ASIC (Application Specific Integrated Circuits) are 800 and 550 times more efficient, respectively, compared with conventional microprocessors for edge detection and stereo matching. The second version is currently being developed in order to further reduce the silicon area size. It uses hybrid analog/digital circuit technology while the first version uses only digital circuits. We are developing a hybrid analog/digital array processor chip which includes a large number of processing elements. Each processing element includes a digital memory unit, a data flow control switch unit and an analog arithmetic/logic unit. The analog arithmetic/logic unit reduces the silicon area size significantly compared with the digital one. The data flows among multiple processing elements in the array chip in a form of analog voltage. The data flow is controlled by the data flow switches. The digital memory unit controls the set-up of the data flow control switch and arithmetic/logic units.

1. INTELLIGENT VEHICLES (MOBILE ROBOTS ON HIGHWAYS)

Intelligent vehicles are mobile robots on highways and are generating rapidly increasing interest. There are two kinds of reasons for the recent increase of interest in this field: social and technical. The social reasons include the fact that conventional approaches for improving highway traffic systems are reaching their limitations and a new approach is necessary. In the past, we expanded the highway systems physically to accommodate ever increasing traffic congestion. It is, however, getting difficult to build additional lanes to existing highways or to make new highways because of environmental concerns, costs and other issues. Intelligent vehicles are expected to be a new approach for developing safe, highly efficient and environmentally-friendly traffic systems. The technical reason is that the underlying technologies, such as signal processing, communication, computers and sensors, have finally reached the level at which the intelligent-vehicle-related devices can be produced at affordable prices [1]. For example, when General Motors and RCA jointly demonstrated an intelligent cruise control system in the 1950s, no one saw any possibilities of converting it into a real product. The system used a microwave radar to measure the three-dimensional (3D) positional relationship between the automated vehicle and the vehicle in front. Steering and speed were automatically controlled so that the automated vehicle followed the vehicle in front. The system was too big and the predicted production cost was out of the question at that time. With significant technological progress made in the related areas, the intelligent cruise control system, for example, is expected to be on market in the near future. Examples of other intelligent vehicle functions are collision warning, collision avoidance, warning for lane changing, obstacle detection for reversing and side collision prediction for exploding side air bags.

2. SENSORS FOR INTELLIGENT VEHICLES

Various sensing approaches are being studied for intelligent vehicles, including vision, microwave, acoustic and laser radar. More research is needed to reach a consensus on what types of sensors are suitable for what intelligent vehicle functions [2]. General merits of the vision approaches are:

(1) Only vision systems have the potential of measuring the lane boundaries without changing the existing roads. Lane sensing is useful for a number of applications. With intelligent cruise control systems, for example, a lane sensing capability makes it possible to measure the distance to the car in front in the same lane, not in the next lane, even on curves.
(2) Since no sensing method can be perfect at this time, it is important that typical drivers can predict and understand when the system might fail. It is an advantage of the vision systems that their characteristics are similar to human visual perception and therefore people can predict easily when they might fail.
(3) Vision systems are passive and do not emit anything. It is not necessary to consider health, regulation and interference issues.

Potential drawbacks of the vision systems are:

(1) The vision systems do not work when human drivers cannot see. It is, however, controversial whether this is a drawback or a merit, because it would be dangerous if people would drive at high speeds in dense fog with overconfidence in the microwave radar. In technical aspects, the functions of the vision systems will be extended by using IR wave ranges.

(2) The history of vision research is shorter than that of microwave radar. This drawback is diminishing with recent increased research efforts.

3. RELATED WORK ON VISION SYSTEMS FOR INTELLIGENT VEHICLES

A number of vision systems have been developed for intelligent vehicles and all of them are still in the research stage [3–5]. This section describes some examples of vehicle applications of vision systems. A popular vision application research for vehicles is road following or lane following. A vision system measures the curvature of the road or lane in front, and controls the steering angle and the vehicle speed to follow the road/lane automatically. A group at Universitat der Bundeswehr Munchen uses six microprocessor boards in parallel for real-time detection of road edges. Each of the six processors is dedicated to a small region in the image frame. Carnegie Mellon University has developed various vision systems for road following, including a system which uses neural networks. Bristol University uses lane markings themselves, instead of lane edges, as explicit objects. The lane markings are extracted based on knowledge of their size, shape and gray-level intensity characteristics. Road models are then used to verify candidates of lane markings. Their road models are simpler than others, such as the one developed at the University of Maryland. The Bristol group assumes that the road surface is on a single straight plane with a single radius.

Other approaches for road/lane detection include a texture-based method by Laboratorire Heudiasyc, and a combination of neural network and texture-based segmentation by Universitat Politecnica de Catalunya. Yamanashi University merges a lit road segment with a shaded road segment using the color characteristics (the percentages of the red and green light vectors in the total light). In this process, the lit and shadowed portions of the road can be merged as the same road region although these portions are different in terms of the intensities. Based on the experimental results of Matsushita Corp., the reliability of the visual recognition of the lanes heavily depends on the weather conditions. For example, the success rates are as high as 97% for day-time and 98% for night-time with fine or cloudy weather, and are reduced to 26% for sunrise/sunset. A group at Mazda sees the necessity of further research in vision systems and knowledge-based reasoning systems for automatic lane following on real highways. They report as much as 5% failure in recognizing lane marks on real highways because of a variety of external disturbances. The knowledge-based reasoning capability is important in deciding actions based on the external information acquired by the vision systems.

A sign recognition system developed at Peugeot classifies road signs into three categories depending on the contour shapes, i.e. octagonal, triangle and round shapes. These shape categories are for stops, danger warnings and less important information, respectively. Both the octagonal stop and triangle danger signs include the color red. The system intends to detect these two categories. A monochrome video camera is installed near a rear view mirror. An optical filter which cuts red light is attached to the camera lens to increase the contrast between the red regions and the white borders in the signs. Closed contours are extracted from the binary edge image and are represented in the Freeman code format. For the classification, they chose a neural network approach over an expert system or a structured programming method because the neural net approach requires a short development time and a short processing time. Experiments were done at medium speeds (40–60 km/h) and signs were recognized two to four times before they were reached.

Intelligent cruise control is an extension of the conventional cruise control system. While the conventional cruise control system controls the engine throttle to keep the vehicle speed at a set speed, the intelligent cruise control system adjusts the set speed depending on the speed of the vehicle in front and other factors. The function to follow a vehicle in front is an important part of the adaptive cruise concept. The car-following system developed at Ruhr-Universität Bochum takes a symmetric object in an image as the back view of the vehicle in front. The system carries out both tracking and identification of the object. The edge image of the object is correlated with deformable 2D models using an elastic net technique.

Daimler-Benz developed a traffic recognition system which takes three steps for recognizing traffic signs. In the first step, color segmentation is carried out using neural networks. The second step generates hypotheses on the image region containing traffic signs and the kind of the signs, based on the color segmentation executed in the first step, priori knowledge on the traffic signs and outdoor scenes. The whole knowledge is stored in a frame-based network. The third and final step evaluates the hypotheses and outputs the result.

Renault's obstacle detection was developed for adaptive cruise and/or collision avoidance. A feature of the system is that it includes two cameras: a usual video camera and a near-IR camera. The system, therefore, has high sensitivity to red lights including tail lights of the car in front. Daihatsu Corp. combined two-camera stereo and optical flow methods for obstacle detection. While the 'obstacle' usually means a slow moving vehicle in front, Universität der Bundeswehr München developed a system to detect vehicles approaching from behind.

4. 3D VISION SYSTEM

The 3D vision systems can be classified into two categories: indirect and direct systems. The indirect systems use single images and calculate distances based on the focus information or other information. The direct systems include time-of-flight and stereo methods. The stereo approaches are based on triangulation and can be classified into three categories: active stereo using laser, passive stereo involving multiple images and optical flow including a time factor. Motion stereo is an example of well-know stereo schemes [6]. It calculates 3D data of objects from a sequence of monocular images. Some of the motion stereo systems belong to the passive stereo and use discrete features such as lines and corners for 3D calculations, while the other stereo methods use optical flow. A potential problem with motion stereo for intelligent vehicles is that we cannot assume that the motion of the camera is known.

A significant problem with binocular stereo vision is a correspondence problem for finding what part in the right image corresponds to what part in the left image. Even an axial layout in which two cameras share the same optical axis cannot solve the correspondence problem without assuming some constraints. One approach to solve this problem is to assume some constraints, like the Marr–Poggio–Grimson algorithm, and the other approach is to use symbolic representation for matching. Binocular stereo vision can be classified into two categories: area-based and feature-based systems [7]. The area-based stereo systems offer the advantage of directly generating a dense disparity map but are sensitive to noise and breakdown where there is a lack of texture or where depth discontinuities occur. The feature-based systems, in contrast, are less sensitive to noise and highly accurate in the depth measurement but provide only a sparse depth map and have problems incorporating the smoothness assumption. Trinocular stereo vision increases the geometric constraints and reduces the influence of heuristic constraints for stereo matching [8, 9].

Mainstream stereo research projects are related to algorithm aspects with less emphasis on real-time processing. The processing time depends on various factors such as complexities of images and models of computers. Some examples of processing times are listed here to show the computation-intensive nature of the stereo algorithms in general; examples of processing times are 174 and 14.5 s [10], 10 min [11], and 1 and 5 h [7]. It is necessary to develop special architectures to make the processing times shorter. Some examples of the processor architectures are described in the following section.

5. VISION SYSTEM ARCHITECTURE FOR REAL-TIME OPERATION

The vision system architectures can be categorized into two groups: ASIC-based and microprocessor-based systems. With the ASIC-based schemes, the major visual processing tasks are carried out by ASIC (Application Specific Integrated Circuit) chips. The microprocessor-based schemes, in contrast, use general-purpose microprocessors as the major components. Many systems combine these two schemes to various degrees. The most significant merit of the ASIC-based methods is that their product designs are highly efficient. They can deliver higher processing speeds, lower power consumption and smaller silicon area sizes, in general, compared with microprocessor-based methods. The ASIC architectures do not have any unnecessary flexibility for the particular applications and therefore much higher processing speeds can be obtained with smaller silicon areas. It is reported that the processing speed was increased by a factor of 2×10^2 to 1×10^3 compared with conventional microprocessor-based approaches [12]. Potential drawbacks of the ASIC-based approaches include high development costs and less flexibility. The development costs are high because the ASIC schemes need to develop dedicated IC (integrated circuits) chips instead of using off-the-shelf microprocessor chips.

6. A DIGITAL ASIC-BASED STEREO VISION SYSTEM

We developed a small, real-time stereo vision system for intelligent cruise control and tried it on highways to identify real problems. The system is based on the digital ASIC approach, and we chose straight-forward and group-force algorithms which are suitable for ASIC implementation in general. Also we put a significant amount of emphasis on choosing the algorithms which match our particular application requirements instead of pursuing the generality of the system.

Figure 1 shows a system block diagram. The total process consists of the following four steps:

(1) Image acquisition: three cameras take images simultaneously.

(2) Feature extraction: features are extracted from three intensity images; the features are vertical edge segments.

(3) Stereo matching: The feature images are shifted to each other for disparity measurements; two binocular stereo, pairs with three cameras, eliminate most false correspondences.

(4) Post-filtering: post-filtering eliminates distance information generated by non-vehicle objects such as lane markings.

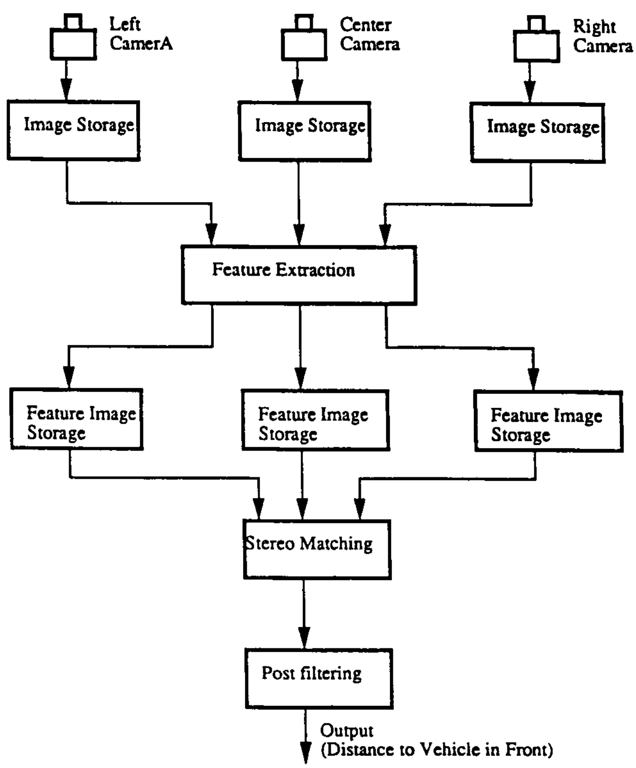


Figure 1. Block-diagram of stereo system.

Figure 2. Front view of test vehicle.

These four steps are described in more details below. For the image acquisition step, three CCD TV cameras are installed at the front of the car as shown in Fig. 2. The right and left cameras are separated from the central camera by 30 cm, respectively. In the second step, positive and negative vertical edges are calculated from the three camera images. Pixels at which the intensity levels increase or decrease significantly from the left to the right of those pixels are defined as positive and negative edges, respectively. We calculate only vertical edges and ignore horizontal edges because the three cameras are located on a single horizontal line. The binary edges are processed by a segment filter in order to eliminate edges which are not part of five-pixel-long vertical edge segments. Only if four or five pixels in the five-pixel-long segment are edge pixels, it is considered a valid edge segment. The vertical segments are used as features for stereo matching.

In the third step, stereo matching is carried out between the right and center images, and between the center and left images in parallel. This dual-matching eliminates a significant portion of the false correspondences. The right and left images are shifted one column by one column to the left and right, respectively. It is considered as a matched trio when three corresponding pixels, each of which is in the right, center and left image, respectively, are all positive or all negative edges. Since the correlation peak of binary edge correlation is very sharp, the edge width of the center image was extended to three pixels while the right and left images have one-pixel-wide edges. Pixels which have matches at multiple disparity values take the nearest distance values for a safety reason. The output of the third step is a distance map which indicates a distance value at every matched pixel. In the following final stage, a histogram is calculated from the distance map image. The horizontal and vertical axes of the histogram are the distance value and the number of pixels which belongs to each distance value. The histogram is then self-convoluted with a window for some distance, e.g. ± 1 shift distance, so that the new histogram represents the number of

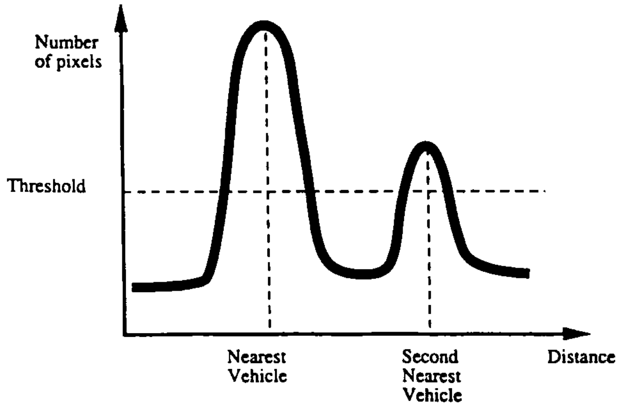


Figure 3. Post-filtering operation.

pixels which belong to each distance range which overlaps each other. The system recognizes the peak as the nearest object if the peak value (i.e. the number of pixels) is larger than the threshold value, as shown in Fig. 3. Through this process, the edge pixels which represent lane markings, for example, are filtered out because they do not make any significant peaks.

7. EVALUATION OF A DIGITAL ASIC-BASED STEREO VISION SYSTEM

At General Motors Research Laboratories, the first version of the ASIC-based processing units for feature extraction and stereo matching were implemented into custom boards (one extended VMS board for each unit) using off-the-shelf CMOS logic chips such as logical-AND chips and counter chips to evaluate the architectures. The speeds of our ASIC-based architectures are compared with microprocessor-based approaches in this section, because high-speed and small size are features of the ASIC-based schemes. The number of operations for feature detection is calculated as follows:

> Number of binary edge detections: 256 × 256 [pixels/image] × 3 [images for right, center, and left] × 10 [times/s] = 2M [pixel-operations/s]

> Each pixel operation includes the following 16 arithmetic/logic operations:
> four additions for calculating two column sums
> two comparisons and two conditional branches for choosing Sobel-ratio or Sobel operator
> one comparison and one division for Sobel-ratio operation
> three comparisons and three conditional jumps for thinning

> Each binary edge requires the following 486 arithmetic/logic operations:
> 324 additions, 81 comparisons and 81 conditional jumps for line segment filtering

If 5% of all the pixels are binary edge pixels, the total number of arithmetic/logic operations required for the feature detection is calculated as follows:

$$2\text{M} \times (16 + 486 \times 5/100) \times 1.5 = 121\text{M arithmetic/logic operations}$$

In the above estimation, a factor of 150% was included to include extra operations such as read-data, write-data, calculate-addresses and other minor operations. Suppose that the machine cycle of a microprocessor is 30 MHz and that every arithmetic/logic operation requires two machine cycles on average, the speed of our system is about eight times faster than a microprocessor-based implementation. The size of our feature detection board is similar to that of an off-the-shelf single board microprocessor. If the board size of our system were reduced by a factor of 100 by replacing low-density small-scale off-the-shelf logic chips with application-specific VLSI chips, the speed/size ratio of the ASIC-based system would be better by a factor of 800.

The following portion of this section will discuss the evaluation of the ASIC-based stereo matching architecture. The application specifications for intelligent cruise control requires the following conditions for the stereo matching unit:

Feature image size:	256×256 pixels
Disparity range:	0–96 pixels
Processing speed:	depth map every 100 ms

The calculation area is 256×256 pixels for disparity-0 (zero) and it is decreased by two for each disparity increment. With disparity-96, the calculation area is $(256 - 2 \times 96) \times 256$ pixels. Two identical series of operations are required: one for positive and the other for negative edge segment images. The total number of matching operations is calculated as follows:

$$((256 \times 256 + (256 - 2 \times 96) \times 256)/2) \text{ [operations/disparity]} \times 97 \text{ [disparities]}$$
$$\times\ 2 \text{ [/stereo\_match]} \times 10 \text{ [stereo\_matches/s]} = 79\text{M [matching operations/s]}$$

If each matching operation requires four machine cycles on average for logical-AND, address-calculation, data-write, data-read and other operations, a speed of 320 MHz is required. This is 11 times the typical 30 MHz speed. Since the board size of our first version is twice as large as that of the typical microprocessor board, the ASIC-based scheme would have a 550 times advantage in terms of the speed/size ratio if the board size of our system were to be reduced by a factor of 100 by replacing the off-the-shelf small-scale arithmetic/logic chips with the ASIC chips.

8. HYBRID ANALOG/DIGITAL STEREO VISION SYSTEM

An emerging technology in the field of vision chips is an analog vision chip scheme. Our group at the Massachusetts Institute of Technology, for example, developed analog chips for image filtering and edge detection, moment extraction to determine object position and orientation, image smoothing and segmentation, depth determination from stereo image pairs, accurate depth determination jointly from imperfect

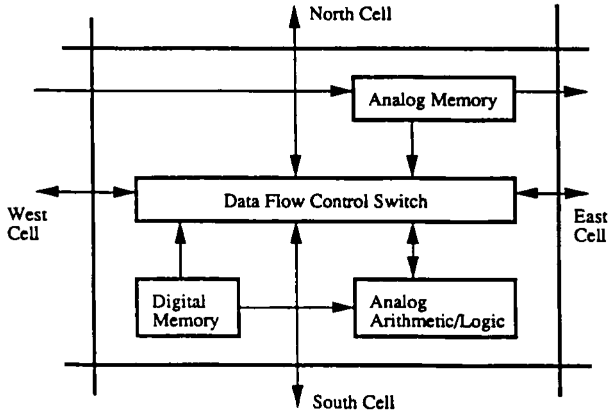


Figure 4. Hybrid analog/digital array processor.

depth and slope data, and camera motion determination, plus other chips to test novel circuit designs or processing methods [13]. Potential merits of the analog vision chips, compared with digital chips, include small silicon areas, high speeds and low mass-production costs.

We are now developing a hybrid analog/digital array processor chip in order to carry out the trinocular stereo operation with a smaller silicon area [14]. The chip has a MIMD (multi-instruction multi-data) scheme, and each processing element consists of a digital memory unit, a data flow control switch unit and an analog arithmetic/logical unit. As shown in Fig. 4, the program in the digital memory unit specifies the set-up of the switches in the data flow control unit and the kind of operations to be carried out by the analog arithmetic/logic unit. The detailed design is underway.

9. CLOSING REMARKS

The first version of the stereo vision system for intelligent cruise control was developed using low-density small-scale off-the-shelf chips at General Motors Research Laboratories. The results of on-highway experiments indicated its acceptable performance. At the Massachusetts Institute of Technology, we are now developing a hybrid analog/digital array processor chip in order to reduce the system's size, cost and power consumption.

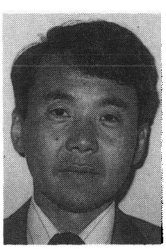
Acknowledgements

The author is indebted to G. G. Dodd, R. B. Tilove and other people at General Motors for developing the first version. Special thanks to the members of the MIT analog vision chip group including B. K. P. Horn, H.-S. Lee, D. A. Martin, C. Sodini, J. White and J. L. Wyatt for the second version which is under development.

REFERENCES

1. Special Report 232, *Advanced Vehicle and Highway Technologies*. Transportation Research Board, National Research Council, 1991.
2. I. Masaki, "ASIC approaches for vision-based vehicle guidance," *IEICE Trans. Electron.*, vol. E76-c, no. 12, pp. 1735–1743, 1993.
3. I. Masaki, *Vision-based Vehicle Guidance*. Berlin: Springer-Verlag, 1991.
4. I. Masaki, ed., *Proc. IEEE Intelligent Symposium '92*, Detroit, 1992.
5. I. Masaki, ed., *Proc. IEEE Intelligent Symposium '92*, Tokyo, 1993.
6. A. P. Tirumalai, B. G. Schunck and R. C. Jain, "Dynamic stereo with self-calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 12, pp. 1184–1189, 1992.
7. P. Cavanagh, "Reconstructing the third dimension: interactions between color, texture, motion, binocular disparity, and shape," *Comp. Vision, Graphics, Image Process.*, vol. 37, pp. 171–195, 1987.
8. N. Ayache and F. Lustman, "Trinocular stereo vision for robotics," in *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 1, pp. 73–84, 1991.
9. M. Yachida, "3-D data acquisition by multiple views," in *Robotics Research: Third Int. Symp.* Cambridge, MA: MIT Press, pp. 11–18, 1986.
10. R. Horaud and T. Skordas, "Stereo correspondence through feature grouping and maximal cliques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 11, pp. 1168–1180, 1989.
11. W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, no. 1, pp. 17–34, 1985.
12. I. Masaki, "Industrial vision systems based on application-specific IC chips," *IEICE Trans. Electron.*, vol. E74, no. 6, pp. 1728–1734, 1991.
13. J. L. Wyatt, Jr, C. Keast, M. Seidel, D. Standley, B. Horn, T. Knight, C. Sodini, H.-S. Lee and T. Poggio, "Analog VLSI systems for image acquisition and fast early vision processing," *Int. J. Comp. Vision*, vol. 8, no. 3, pp. 217–230, 1992.
14. I. Masaki, "New architecture paradigms for analog VLSI chips," *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits*. IEEE Computer Science Press, in press.

ABOUT THE AUTHOR



Ichiro Masaki was born in Japan in 1948. He earned his BS and MS from Waseda University and PhD from Osaka University. He has previously been working at General Motors Research Laboratories (Michigan, USA), Unimation (Connecticut, USA) and Kawasaki Heavy Industries (Tokyo and Kobe, Japan). He is currently a Research Scientist at Massachusetts Institute of Technology. His research interests include vision systems, custom chip architecture, robots and intelligent vehicles.