AN IMAGE RECOGNITION SoC “Visconti™” FOR AUTOMOTIVE APPLICATIONS

Ryuzo Okada[1)], *Hiroaki Nakai*[1)], *Kenji Furukawa*[1)], *Tatsuo Kozakaya*[1)], *Yasuhiro Taniguchi*[1)], *Jun Tanabe*[2)], *Takashi Miyamori*[2)], *and Tatsuya Shimoike*[3)]

1) Corporate R&D Center, Toshiba Corporation
2) SoC R&D Center, Semiconductor Company, Toshiba Corporation
3) Automotive System Division, Toshiba Corporation

Email: ryuzo.okada@toshiba.co.jp

ABSTRACT

In this paper, we present an image recognition SoC named “Visconti™” [1], which has been developed to provide advanced safety assistance for automobile drivers. This SoC is an 18-GOPS multi-VLIW processor, in which three 3-way VLIW processors and peripheral modules such as memory controllers, video I/Os and an affine transformation module are integrated. The SoC design is based on a configurable MeP (Media embedded Processor) architecture [2, 3], which supports superior capabilities for automobile image processing, i.e. high compute performance, low cost and low power dissipation. We have implemented several types of our previously proposed image recognition algorithms on the SoC, and describe three example applications briefly in this paper.

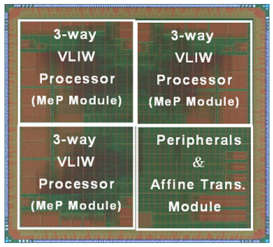
1. INTRODUCTION

In addition to radar, vision sensing using vehicle-mounted cameras have recently emerged as an important technology for providing advanced safety assistance to automobile drivers, e.g. lane departure warning, obstacle detection for vehicle navigation, in-vehicle driver/passenger monitoring. Low-cost high-performance processors for real-time vision systems are therefore strongly demanded for implementation. Several semiconductor companies have developed high-performance image processing LSIs aiming at mass production (vision instruction processor [4], IMAP-CE [5]) but few are currently available for automotive use.

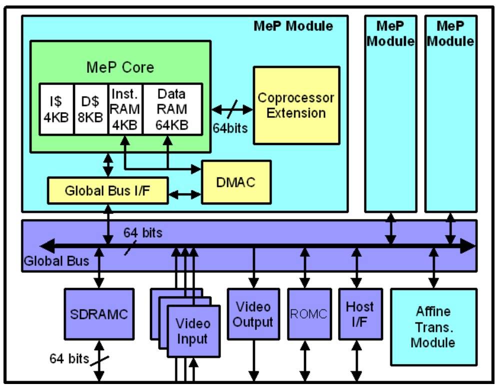
In our previous study [6], we have investigated processor performance necessary for image recognition in automotive use (obstacle detection) and reported that at least a GHz-class processor with SIMD (Single Instruction Stream, Multiple Data Stream) instruction support is necessary although the processor must have small die size, i.e. low cost, as well as robustness under the severe automotive environment, such as wide temperature range and vibration. According to the investigation, we designed a new image recognition processor named ”Visconti™” (VIsion based Sensing, CONTrol, and Intelligence) which possesses an efficient architecture for processing a variety of image recognition algorithms while fulfilling requirements for automotive use. The strategy we adopted in the design is to incorporate multi-level parallelism inherently existing in image processing algorithms into the architectural structure. More specifically, instead of a single high-performance microprocessor, we integrated three 3-way VLIW (Very Long Instruction Word) processors which executes SIMD instructions and a hardware module for image transformation in a single SoC. Thus, we can easily utilize instruction level parallelism by the VLIW architecture and data level parallelism by the SIMD coprocessor extension. Furthermore, task (thread) level parallelism can be utilized by its CMP (on-Chip Multi-Processor) architecture. This multi-grain parallel architecture can effectively provide different levels of parallelism which can meet a variety of user’s target applications.

We have implemented several image recognition algorithms on the SoC [7, 8, 9, 10] in order to realize applications for providing diverse safety support to automobile drivers, such as obstacle surveillance using three single cameras, front obstacle detection using stereo cameras, and face recognition for driver identification. The algorithms run in real time on the SoC by making effective use of the multi-grain parallel architecture, whereas they require a GHz-class CPU for real-time execution on an ordinary PC.

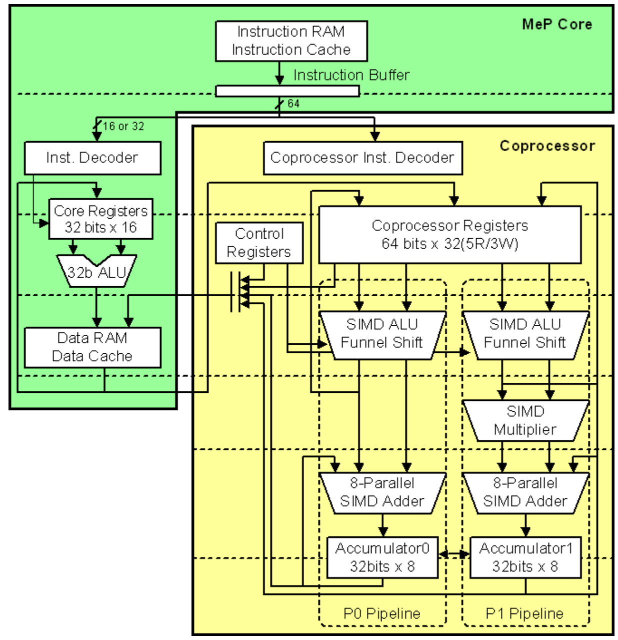
In the following sections, we firstly describe the architecture of Visconti™, and then briefly describe the three image recognition applications for automotive use.



(a) Micrograph



(b) Block diagram



(c) Structure of a MeP core and coprocessor

Fig. 1. Image recognition SoC, Visconti™. Visconti™ consists of three processing modules (MeP modules), an affine transformation module, and peripherals. Each of the MeP modules contains a VLIW processor with the customized VLIW coprocessor extension having two 64-bit SIMD pipelined datapaths to organize a 3-way VLIW processor.

2. THE ARCHITECTURE OF Visconti™

In this section, we describe the architecture of Visconti™. The SoC, an 18-GOPS multi-VLIW processor, is designed to ensure efficient performance for the computations commonly dominant in image recognition algorithms, which will be described in the following sections, while satisfying requirements for automotive use: e.g. operating temperature range of -40 to +85 degrees Celsius.

Figure 1(a) and 1(b) show the micrograph and the block diagram of the SoC consisting of three processing modules (MeP modules) for task (thread) level parallelism, an affine transformation module for accelerating data transformation and peripherals. Each of the MeP modules contains a VLIW processor designed based on Toshiba's configurable processor architecture (MeP core) [2, 3] with the customized VLIW coprocessor extension to organize a 3-way VLIW processor (see Figure 1(c)) for providing instruction level parallelism. The coprocessor has two 64-bit SIMD pipelined datapaths, which provide data level parallelism. One of the datapaths consists of an 8-parallel SIMD ALU and an accumulator. The other has an 8-parallel SIMD MAC in addition to an 8-parallel SIMD ALU. Several types of SIMD operations, e.g. convolution, accumulation, pixel shift (funnel shift), are supported in the instruction set of the coprocessor for the computations often required in image processing algorithms, such as image filtering, block matchinga and vector inner product calculation. Each MeP module contains 4KB instruction cache, 4KB instruction RAM, 8KB data cache, and 64KB data RAM with a DMA controller (DMAC) internally so that memory access latency can be hidden by double-buffering data transfer.

The SoC also contains a special processing module (affine transformation module) for data transformation, e.g. affine transformation of an image, which is difficult to be accelerated by SIMD operations because of discontinuous memory access. Although this module is named after affine transfor-

Table 1. Specifications of Visconti™

Feature	Specification
Technology	0.13 μm CMOS 6-layer metal
Peak performance	18 GOPS (6 GOPS$\times$3 processors)
Clock frequency	150 MHz
No. of transistors	21 million (17 million for memory)
On-chip memory	260 Kbytes
Power dissipation	1 W at 1.5 V
Operating Temp.	$-40°$C to $+85°$C
Chip size	6.98mm $\times$ 6.98mm (48.7mm^2)
Package	456 pin PBGA

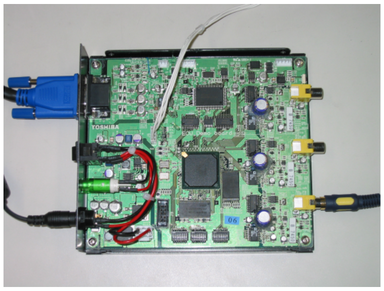


Fig. 2. A prototype electronic control unit (ECU). The ECU consists of a Visconti™ SoC in combination with a few peripherals such as SDRAM memory chips, a power supply and video input channels.

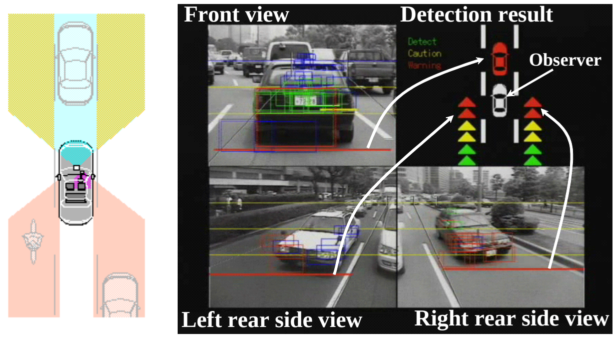


Fig. 3. Obstacle surveillance system using three cameras. The cameras are positioned to observe the front, the right rear side and the left rear side of the driver's car (left). The result of obstacle detection (right) shows that the system can detect preceding cars and passing cars simultaneously.

mation of an image, which is a typical type of data transformation, the module is capable of performing any types of data conversion by configuring a conversion table that defines the conversion of each datum.

Several peripherals, e.g. an SDRAM controller, three video inputs, a video output, a host interface and a Flash ROM controller are also integrated into the SoC for the convenience in building practical application systems.

The specifications of the SoC are listed in Table 1. The SoC operates at the frequency of 150MHz in the worst case with the operating temperature range of -40 to +85 degrees Celsius (-40 to $+185$ degrees Fahrenheit). The power dissipation is approximately 1W at the peak performance of 18 GOPS. These specifications fulfill the requirements for executing sophisticated image recognition algorithms, which is described in the next section, for automotive use.

3. AUTOMOTIVE APPLICATIONS USING Visconti™

Figure 2 shows a prototype of an electronic control unit (ECU) for automobiles using a Visconti™ SoC in combination with a few peripherals such as memory chips, a power supply and video input channels. It is easy to construct a vision system using the ECU and vehicle-mounted cameras, which provides various image sensing and recognition functions, e.g. an obstacle detection system that alerts the driver in the event of dangerous situations.

In this section, two applications of obstacle detection using the prototype ECU are briefly described, which are (1) obstacle surveillance using three cameras positioned to observe the front as well as the right and left rear side for detecting preceding cars and passing cars simultaneously, and (2) a front obstacle detection with lane detection using a pair of stereo cameras. We have also implemented a real-time face recognition system using the same prototype ECU, which can be applied to anti-theft protection by driver identification. This application shows that Visconti™ is also capable of statistical computations required for accurate face recognition, which are quite different from the ones in obstacle detection. Thus, using the same hardware but replacing the software, we can realize various image recognition systems for automotive use.

3.1. A surveillance system using three cameras

We have developed a surveillance system [11, 7] for automobiles using Visconti™ and three vehicle-mounted cameras. The cameras are positioned to observe the front, the right rear side and the left rear side (Figure 3 left) so that the system can detect approaching cars in all directions at the same instant (Figure 3 right). For such a warning system, reducing false alerts is very important so as to ensure that the system is neither annoying nor causes accidents by the false alerts. Thus obstacles (cars) must be detected distinctively from non-obstructive objects, i.e. texture on the road surface such as road signs, shadow, and so on. We use a ”motion constraint” between horizontal line segments observed and tracked in each image from a single camera for obstacle/non-obstacle discrimination. The principle utilized in the method can be simply explained by Figure 4. The ratio of two intervals, which are vertical distances between the three horizontal lines, is constant if three horizontal lines belong to an obstacle (Figure 4 left). On the other hand, the lower interval becomes wider than the upper interval and the ratio of the two intervals is not constant for three horizontal lines belonging to the ground plane, i.e. road surface (Figure 4 right). For accurate discrimination, the horizontal lines on the obstacle are selectively detected by calculating “cross ratios” using the horizontal lines and vanishing lines of the ground plane and that of the plane representing obstacles (see [7] for details).

The process flow of the obstacle detection consists of

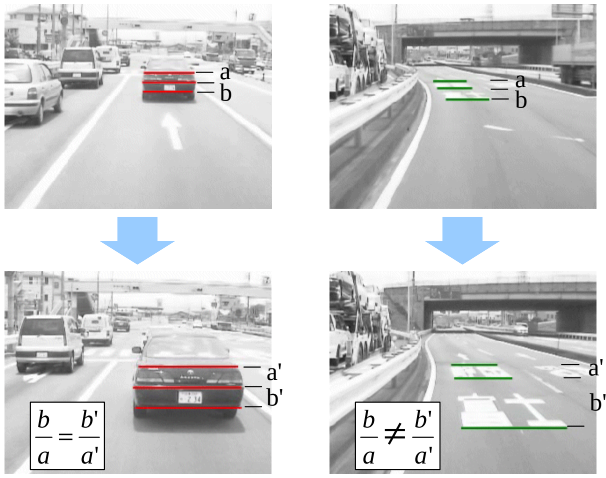


Fig. 4. Obstacle detection by interval ratio comparison. The interval ratio is constant for the three horizontal lines belonging to an obstacle (left), whereas that for the three lines on the road is not constant (right).

Table 2. Processing time [μs] for the surveillance system.

Major processing components	Before optimization	After optimization
Calculating OC	10162	3884
Detecting HLSs	74849	2697
Tracking HLSs	101369	2793
Obstacle discrimination	7694	1232
Total (1 frame)	232058	14956

the following steps; 1) detection of horizontal line segments (HLSs), 2) tracking of the HLSs in an image sequence, and 3) obstacle discrimination by cross ratio calculation. Almost the same steps are implemented in all MeP modules working independently. The dominant computation in the step 1) is calculation of separability [12] of image intensities, and that in the step 2) is template matching using orientation code matching (OCM) [13]. The OCM is a variation of template matching, where the similarity measure between any two images is defined based on the quantized orientation of the edges (orientation code; OC). We implemented these calculations using SIMD instructions and the affine transformation module, and evaluated the efficiency of our implementation. Table 2 shows typical computational time per frame for QVGA images for major processing components with and without optimization to the Visconti™ architecture. Template matching and separability calculation are greatly accelerated because two SIMD operations are executed by the coprocessor in parallel, so that the computational times of "Tracking HLSs" and "Detecting HLSs" after the optimization are about 40 and 30 times faster than those before the optimization, respectively. We computes the orientation codes by the data transformation function of the affine transformation module, and "Calculating OCs" is accelerated about three times. The overall processing rate of the system is about 10-50 [ms/frame] (depends on the number of obstacles).

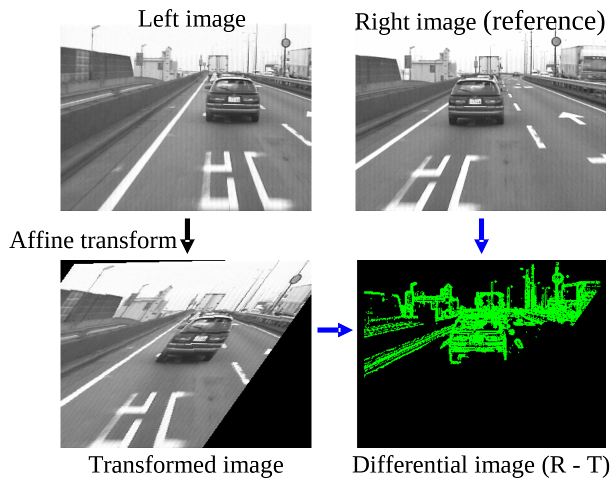


Fig. 5. Front obstacle detection based on planar-projection stereo

3.2. Front obstacle detection using stereo cameras

A front obstacle detection system using a pair of stereo cameras has also been developed using the same Visconti™ platform [8]. In this system, the cameras are used to measure the distance to an obstacle and also to ensure robust obstacle detection under noisy conditions, e.g. caused by weather or poor illumination, by exploiting geometric relation between cameras. The processing flow of the method consists of the following three steps; 1) lane detection, 2) obstacle detection based on planar transform, and 3) obstacle verification by region matching.

First, lane markings in the image from the camera on the right side are detected based on edge segment detection. The image is also transformed to the plan view in order to detect correct pair of lane markings using the parallelism. The lane marking detection provides useful information for lane departure warning and also contributes in reducing computational cost by focusing on the driving lane.

The obstacle detection in the step 2) is based on the assumption that a vehicle moves on a flat plane such as a road, so that the algorithm for obstacle detection can be significantly simplified in comparison with conventional stereo matching approaches [9]. The obstacle detection basically consists of the following three sub-steps (see Figure 5); i) let the image from the camera on the right side be a reference one, ii) then the image from the camera on the left is

Fig. 6. Results of front obstacle detection in a heavily raining scene and a night scene.

transformed to the reference view assuming that all image points arise from the ground plane (GP), iii) comparing the reference image and the transformed one. The points arising from the ground plane have no disparity between these images, while those on obstacle areas have disparities due to the height form the ground plane. Therefore, simple image subtraction enables us to detect obstacle regions that have non-zero image difference and it is unnecessary to search for corresponding points. The image transformation in ii) is a planar transformation between a pair of images and is defined as an affine transformation.

Figure 6 shows the examples of the obstacle detection under bad weather conditions. Obstacles are correctly detected in the images, and the results illustrate the capability of the proposed method. Unfortunately, there could be a possibility of false detection in the obstacle detection process if the assumption that the vehicle moves within a horizontal plane is violated; e.g. the texture of the road surface would be falsely detected as an obstacle. This can occur when the geometric parameters of the stereo cameras become inappropriate due to road inclination and camera movements and the road inclination changes at the mid point of the read. To cope with this drawback, we have developed an efficient method for obstacle verification method including stereo parameter update. It is based on the planar transformation and simple block matching (see [8] for the details).

In the preliminary experiment, at least a 1.7-GHz microprocessor for generic PCs with SIMD instructions was necessary for the real-time processing of the obstacle detection described in this section. We implemented all of the image processing techniques on the Visconti™ platform, which is capable of executing all processes at a rate of 30 frames/sec (QVGA).

3.3. Compact face recognition system

Currently, personal identification using biometric information attracts a great deal of interest in the area of security. Face recognition is superior to the alternatives in terms of usability, being contact-free and intuitive for users. However, it has not been available in popular purposes, e.g. automobiles, doors, robot vision and mobile phones, due to two main problems: dealing with variations of face patterns and large, expensive devices, e.g. high performance CPUs. We can solve the problems by constructing a compact and low-cost system [14] using Visconti™ with a very robust recognition method [10] (see Figure 7).

Fig. 7. Typical result of face recognition. The upper left image is an input from the camera with the extracted face feature points. The small upper right image indicates the result of face recognition.

The face recognition in the proposed system consists of two main steps; 1) face detection [15] and 2) face identification. In the first step, a face position is located in video input. Eyeballs and nostrils, i.e. face features which have circular shapes, are found by an image filter called a separability filter [12] by means of the degree of separation between inner and outer regions of a circle (see Figure 8). The detected features are verified by matching with the eye and nose dictionary patterns to ensure the detection. We also use geometric relationship between the features to select a correct set of them. The direction and size of the face pattern are normalized using the affine transform based on the retrieved four points. The extracted face patterns are shown in Figure 8.

The second step identifies the located face pattern using a statistical pattern recognition technique named "constrained mutual subspace method (CMSM)". The method has two main advantages in realizing accurate and robust recognition; 1) similarities between input and registered face patterns are calculated using a video stream, i.e. multiple face images, to cope with variations in facial directions and expressions, 2) "constrained subspace" which is orthogonal to individual facial patterns (subspaces) is introduced in similarity calculation to cope with variations in lighting conditions (see [10] for details). The method has been incorporated into a commercial access control system [16].

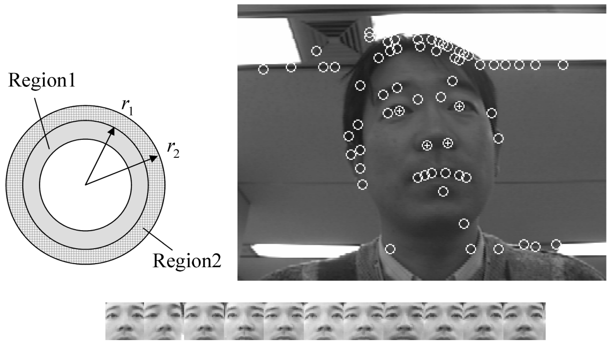


Fig. 8. Face detection. Separability filter for detecting circular feature points (top left), detected face feature points (top right), and cropped face patterns (bottom). Detected face feature points are shown by white circles, and a correct set of face feature points is shown by the circles with '+' marks inside.

The implementation of the face recognition is optimized to make the best use of the hardware features for real-time operation, e.g. integerization and fixed-point representation of the recognition algorithm, dynamic allocation of different tasks to MeP modules. As the result, the system operates at 20 frames/sec with keeping high recognition accuracy, 99.59%, which is comparable to that of a state-of-the-art system. Thus, the system is very compact and operates at low power, making it suitable for many purposes, including driver identification for automotive use.

4. SUMMARY

We have introduced a SoC named "Visconti™", which has an efficient architecture for embedded applications of image recognition technologies. We have also demonstrated that it is easy to realize a variety of vision systems using the same hardware but modifying the software. Owing to its outstanding capabilities, we expect Visconti™ to be used for various image recognition applications in automotive systems.

5. REFERENCES

[1] J. Tanabe et al., "Visconti: Multi-VLIW image recognition processor based on configurable processor," in *Proc. of IEEE Custom Integrated Circuits Conference*, 2003, pp. 185–188.

[2] T. Miyamori, "A configurable and extensible media processor," in *Proc. of Embedded Processor Forum*, 2002.

[3] "Media embedded processor," http://www. MePcore. com.

[4] W. Raab et al., "A 100-GOPS programmable processor for vehicle vision systems," *IEEE Design & Test of Computers*, pp. 8–15, 2003.

[5] S. Kyo et al., "A 51.2 GOPS scalable video recognition processor for intelligent cruise control based on a linear array of 128 4-way VLIW processing elements," in *Proc. of IEEE International Solid-State Circuits Conference*, 2003, pp. 48–49.

[6] Y. Kondo et al., "A 4GOPS 3way-VLIW image recognition processor based on a configurable media-processor," in *Digest of technical papers of IEEE International Solid-State Circuits Conference*, 2001, pp. 148–149.

[7] R. Okada et al., "Obstacle detection using projective invariant and vanishing lines," in *Proc. of IEEE International Conference on Computer Vision*, 2003, pp. 330–337.

[8] H. Nakai et al., "A practical stereo scheme for obstacle detection in automotive use," in *Proc. of IAPR International Conference on Pattern Recognition*, 2001, vol. 3, pp. 346–350.

[9] H. Hattori and A. Maki, "Stereo without depth search and metric calibration," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 177–184.

[10] K. Fukui, "Face recognition using multi-viewpoint patterns for robot vision," in *Proc. of International Symposium of Robotics Research*, 2003.

[11] K. Furukawa et al., "Onboard surveillance system for automobiles using image processing LSI," in *Proc. of IEEE Intelligent Vehicles Symposium*, 2004, pp. 555–559.

[12] K. Fukui, "Edge extraction method based on separablity of image features," *IEICE Trans. Inf. & Syst.*, vol. E-78-D, no. 12, pp. 1533–1538, 1995.

[13] F. Ullah, S. Kaneko, and S. Igarashi, "Orientation code matching for robust object search," *IEICE Trans. on Inf. & Syst.*, vol. E84-D, no. 8, pp. 999–1006, 2001.

[14] T. Kozakaya and H. Nakai, "Development of a face recognition system on an image processing LSI chip," in *Proc. of CVPR Workshop on Face Processing in Video*, 2004.

[15] K. Fukui and O. Yamaguchi, "Facial feature point extraction method based on combination of shape extraction and pattern matching," *Systems and Computers in Japan*, vol. 29, no. 6, pp. 49–58, 1998.

[16] "FacePass," http://www3.toshiba.co.jp/snis/scd/fps/.