15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP

A SPEED ADAPTIVE EGO-MOTION DETECTION SYSTEM USING EDGE-HISTOGRAMS PRODUCED BY VARIABLE GRADUATION METHOD

Jia Hao and Tadashi Shibata

Department of Frontier Informatics, School of Frontier Science
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
haojia@if.t.u-tokyo.ac.jp, shibata@ee.t.u-tokyo.ac.jp

ABSTRACT

An ego-motion detection system has been developed, where edge features are extracted from an image and analyzed to detect any motion from the camera holder instead of using the conventional method of comparing pixel intensities. Such edge-based feature representation scheme reduces the computational complexity and increases the accuracy, thus being better suited to hardware implementation due to its simplicity. In this paper we have enhanced the reliability and flexibility of the system by introducing a new pre-processing scheme in edge detection and an automatic speed adaptation capability in local motion detection. The pre-processing improves the local motion detection accuracy by only highlighting the apparent general contour edges, while filtering out insignificant features in the background which may lead to misjudgement. The automatic speed adaptation capability improves the system and renders it more flexible to accommodate to more complex motion patterns with variable speeds. The system performance has been demonstrated by simulation experiments and the robustness against disturbing moving objects in the scene has also been shown.

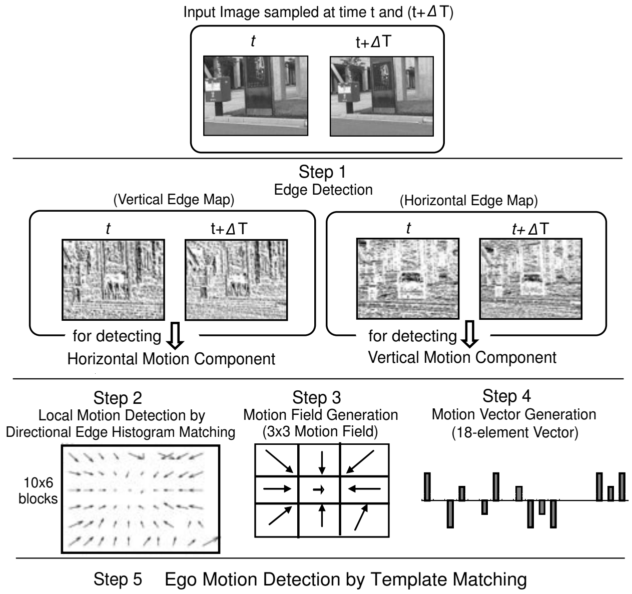


Figure 1: Flowchart of ego-motion detection.

1. INTRODUCTION

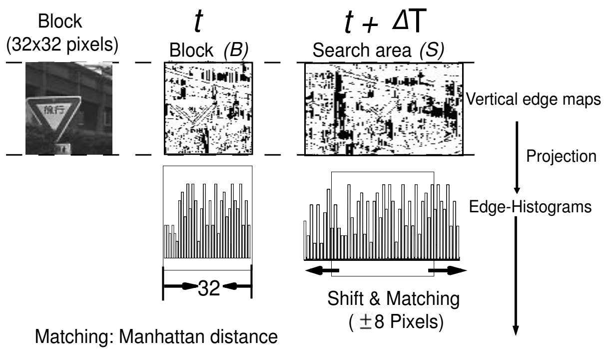
For human being, motion recognition is an essential ability in understanding the environment for survival. Therefore a number of studies have been carried out to understand the mechanisms of motion recognition. The motion recognition is usually considered to be composed of two steps, i.e., *motion detection* and *motion analysis*. Motion detection is a technique to detect the movement of a specific object using visual information, and a number of algorithms have been proposed in this area[1, 2, 3, 4, 5, 6].

Among these studies, ego-motion detection is one of the hottest topics used in such applications as automatic vehicle navigation, robot control, and security surveillance. Being different from general motion detection which aims at tracking a certain object, ego-motion detection identifies the global motion path of an observer by analyzing the information in its visual field. However, most of the existing algorithms found in literatures estimate the motion models by solving some complex equations with floating-point calculations. Therefore they are computationally very expensive and not compatible to real time applications.

In the previous work [7], a hardware friendly algorithm for ego-motion detection was proposed. All operations are executed with 1-bit or integer operations, i.e. in fixed-point calculation, thus being easy to implement with compact hardware circuitry, which is crucial in building real time systems.

However, in the previous system, the reliability and flexibility are not sufficient and improvements in the performance have been awaited. For example, in local motion detection, insignificant features abundant in the background sometimes confuse local motion detection, resulting in the misjudgement of the system in certain cases. The other problem is that the system has been adapted to constant-speed motions, thus being not compatible to irregular motions including abrupt motion changes.

In this paper, we have introduced two new schemes to the system, namely the *variable graduation method* and *automatic speed adaptation scheme* in order to resolve these problems. In the former, it is aimed to promote the advantage of the edge-based feature representation to larger extent. Only the salient contour in the edge map is retained and emphasized by discarding insignificant details in the background. The basic idea of the speed adaptive scheme is an



$$D(B,S) = \sum_{i=1}^{32} | B(i) - S(i) |$$

Figure 2: Directional Edge Histogram Matching (DEHM) utilized for detecting horizontal motion component of the image in a 32x32-pixel block. Edge maps are taken from image pairs at time t and $t+\Delta T$. Then histograms are generated by projection of edge flags. The best matching in the search area yields the horizontal motion component.

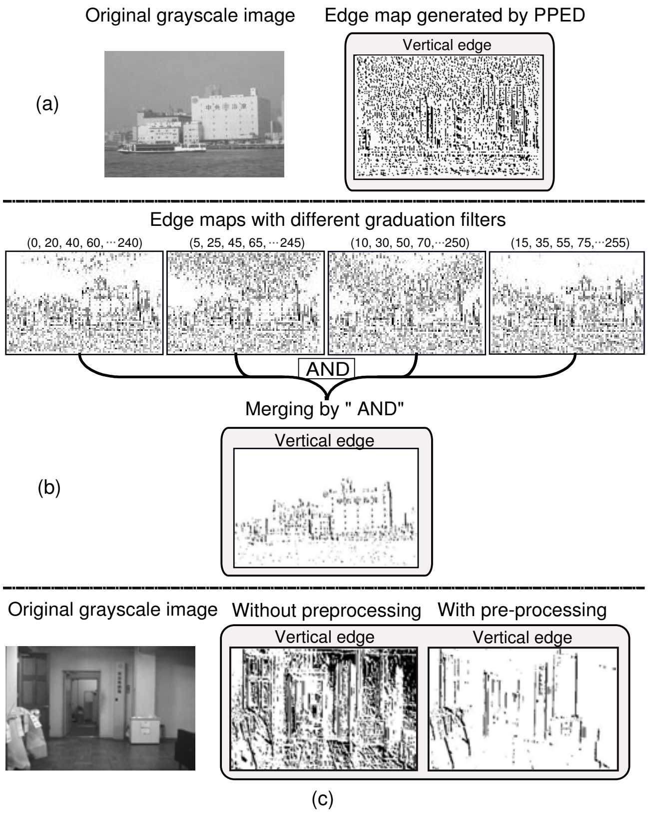


Figure 3: Procedure of variable graduation method. Original image and edge map without pre-processing (a). Generation of vertical edge map with pre-processing (b). Edge detection is operated four times with separate parallel graduation filters. Final edge map is generated by 'AND' operation. Performance comparison for an indoor scene (c).

automatic adjustment of the frame interval that determines the accuracy in local motion detection. This has improved the quality of global motion representations, thus enhancing the accuracy of the following template matching step to determine ego-motion patterns.

The paper is organized as in the following. Firstly, in Section 2, the proposed ego-motion detection system is presented referring to the previous work, and the two new schemes introduced in the present work are described. Then, the simulation experiments and the results of performance evaluation are presented in Section 3, and conclusions are given in Section 4.

2. EGO-MOTION DETECTION BASED ON EDGE HISTOGRAM MATCHING

2.1 System overview

The processing flow in the previous system is shown in Fig. 1, which is carried out in five steps.

Step 1: Input data are given as a sequence of motion pictures. Horizontal and vertical edges are extracted from each frame according to Projected Principal-Edge Distribution (PPED) scheme [8, 9, 10].

Step 2: As the result of Step 1, horizontal and vertical edge maps are generated. Then, two edge maps at t and $t+\Delta T$ are utilized to detect local motions based on Directional Edge Histogram Matching (DEHM) scheme, which is explained in detail in Fig. 2. Local motions are detected in the array of 10x6 local blocks.

Step 3: Local motions are grouped into nine separate areas in a 3x3 form. Local motions within each area are summarized and an area motion is determined. The set of motions in the array of 3x3 areas represents a motion field.

Step 4: Then an 18-element motion vector is generated by collecting x and y components of the nine area motions.

Step 5: Finally, ego-motion detection is carried out by template matching. Four types of motions, i.e., moving vertically, moving horizontally, zooming, and rotating are provided as motion patterns and one of them is selected.

2.2 Variable intensity graduation image pre-processing for highlighting contour edges

In this work, it is aimed to detect the global motion tendency in the visual field, rather than the details. For example, an image sequence of a galloping horse with its hair flowing in the wind, the information contained in the areas displaying flowing hair is of no use to determine the global motion, but can have a negative impact. Humans understand the environment by only relying on salient features in the scene. For example, seamen can recognize a coastline from very little visual information, although the field of view is dominated by monotonous water and sky. Most of the information is contained in the boundary regions between the sky and the water, or the water and the land. Namely, apparent contour features present in the scene play an essential role. In order to retain only essential contour features while discarding superfluous information, we have introduced *variable graduation method*, which is explained in the following.

The original grayscale image and its vertical edge map generated following the PPED scheme [8, 9, 10] are shown in Fig. 3(a). Numerous edge flags are detected in the region of water and sky, which include the edges arising from camera noises, JPEG block noises, etc. The reason for such a numerous edge detection is due to the PPED algorithm which has been designed and adapted to detect all delicate edge information from a grayscale image in order to use them in medical radiograph analysis and diagnosis. Certainly such edges degrade the performance of local motion detection by DEHM. To overcome this problem, a variable graduation pre-processing scheme, applied before edge detection, is proposed for highlighting only the most significant features. The algorithm is explained in Fig. 3(b).

By gradually reducing the graduation, less significant features are filtered out and only the essential general contour edges are retained, and then used for motion detection. In this process, pixels with similar intensities are converged into the same group. We reduce the graduation of image intensity to filter out unimportant details.

Although image is smoothed in the operation, the prominent edge features are left, while the areas without obvious variance of intensity are flattened. In the present work, the full-scale 8b intensity graduation (0, 1, 2, $\cdots$ 255) has been reduced to (0, 20, 40, $\cdots$ 240), for instance. Here, 20, 40, etc. represent the intensity boundaries in the reduced graduation. In Fig. 3(b), four vertical edge maps generated from the original image in Fig. 3(a) with four different intensity graduation boundaries are shown. Then, 'AND' operation is carried out on these four edge maps. Namely, edge flags are retained only at the pixel sites in which edge flags are all present in four maps. The resultant edge map is also shown in Fig. 3(b) as the convergence result. It is interesting to observe that all superfluous edges are filtered out, and only essential contour features are retained.

To confirm the effectiveness of the pre-processing, another experiment was carried out for an image sequence of an indoor scene in Fig. 3(c). As shown in the figure, the contours of building structures come out more clearly with the pre-processing scheme.

2.3 Automatic speed adaptive scheme

In the previous ego-motion detection system, two frames of images sampled at two different times are used for operating motion detection. This algorithm has achieved a high level of accuracy when the camera is moving at a uniform speed.

When the camera moves with a variable speed, the constant interval of sampling image pairs result in a different amount of motion. Therefore, considering an actual egoemotion in which speed variation is inevitable, a stationary interval setting in the previous algorithm will lead to incorrect results in local motion detection.

In local motion detection by DEHM shown in Fig. 2, search area is located ± 8 pixels offset from the original block

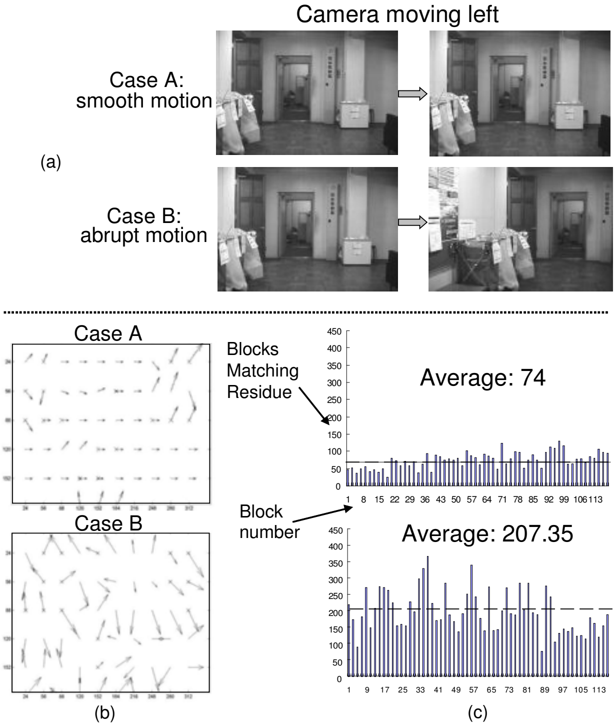


Figure 4: Image sequences of smooth motion (Case A) and abrupt motion (Case B) (a) and resultant motion fields (b). Distributions of matching residues for Case A and Case B (c).

location.

When a sudden large motion is encountered not fitting to the frame interval of image pair sampling, the object motion is out of the ± 8 pixels range and erroneous local motion detection occurs. However, such an unexpected abrupt motion can be easily eliminated using the technique illustrated in Fig. 4.

Two types of camera motions are illustrated in Fig. 4(a). A smooth camera motion in Case A and an abrupt camera motion in Case B. Motion fields generated from these two motion types are also given in Fig. 4(b). In Fig. 4(c) are given block matching residues for Case A and Case B. The matching residues are the Manhattan distances yielding the best match of edge histograms in local motion detection.

If correct local motions are detected from the smooth motion, their residue average is small. But if an abrupt motion occurs as in Case B, the average value of the matching residue increases. By setting the threshold for the average residue at 120, such abrupt motions can be eliminated very effectively.

A speed-adaptive scheme has been introduced to the algorithm to make the system more flexible and capable of analyzing more complex motion patterns. The procedure is explained in Fig. 5. The camera's global displacement is determined by calculating movement of the five largest motion areas, while the frame interval is fixed at a constant value. Once the camera's global displacement has reached the pre-

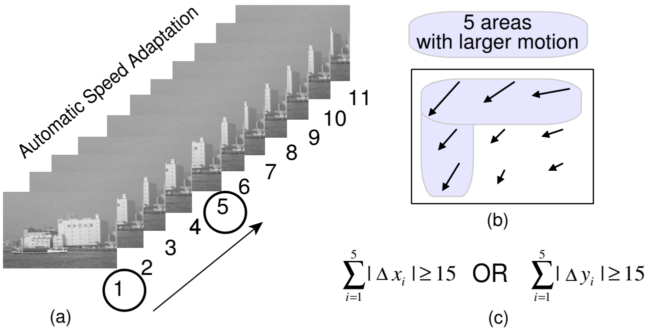


Figure 5: Automatic speed adaptation procedure. (a) Temporal image sequence used for motion vector generation. (b) Camera's global displacement is determined as the average of five largest area motions. (c) Threshold condition to determine the timing of motion vector generation.

determined threshold, the motion field is used to generate a motion field vector. In our simulation the threshold of 15 pixels was employed.

3. EXPERIMENTAL RESULTS

Figure 6 shows the local motion vectors detected from an image sequence taken by a camera on a boat with (a) or without (b) the image pre-processing phase using the variable graduation method. Without any pre-processing, erratic local motions appear due to some superfluous edges coming from the background. However, with the pre-processing phase, most of the background noise is eliminated, resulting in more consistent local motion vectors. As a result, it is determined that the observer is moving to upper-left, the correct result due to the floating motion of the boat.

The reason for including a speed adaptive scheme is to efficiently extract and make use of the motion information from the video sequence. It optimizes the procedure for selecting frame pairs used to detect motion, thus reducing the probability of producing invalid motion vectors. Figure 7 shows the advantage of using this technique. Figure 7(a) shows the result of speed adaptation for the frame sequence shown in Fig. 5(a). As the frame sequence progresses, the adapted frame interval changes a lot. It effectively reveals the irregular motion speed. In Figs. 7(b) and (c), the result of motion detection without or with the speed adaptation scheme are shown, respectively. When the speed adaptation is not executed, the result includes more than 60% invalid motion vectors produced due to abrupt motions. On the other hand, with the speed adaptation, the occurrence of invalid motion detection is effectively suppressed.

Also the singular result is displayed as a motion field in the figure. When the ship is moving diagonally due to the floating boat motion, motion is explained as moving down correctly, while it is interpreted as rotation without speed adaptation.

The performance of the system against disturbing moving object is given in Fig. 8. The effect of another ship

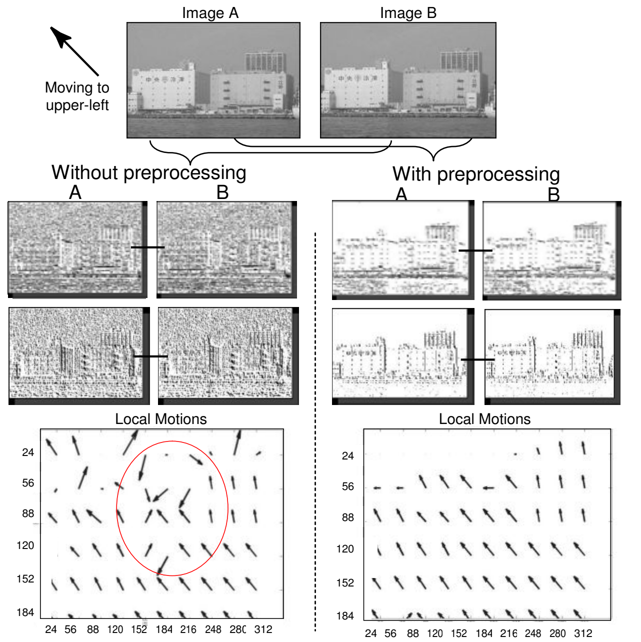


Figure 6: Illustration of performance of pre-processing. Edge maps and resultant motion fields detected with or without pre-processing are shown.

moving in the foreground is shown. The obstructing ship is passed over the sequence and the scale of the obstruction was varied from 0 to 1. The scale of 1 means the height of the ship image is equal to the height of the original image sequence. It is seen that the accuracy decreases drastically when the scale of the disturbance overrides 50% of the background. The reason for this is that only the sky and sea which include few features are left for detecting motions. However, such an illusion is also very likely to happen for human observers.

4. CONCLUSIONS

The performance of the ego-motion detection system proposed in [7], based on directional edge histogram matching, has been enhanced by introducing two new schemes. First a new pre-processing phase prior to edge detection and secondly an automatic speed adaptive processing phase. The pre-processing improves the local motion detection accuracy by only highlighting the apparent general contour edges, while filtering out insignificant features in the background which may lead to misjudgement. The automatic speed adaptation capability makes the system more flexible when interpreting more complicated motion behavior. The system performance has been verified by simulation experiments and the robustness against disturbances caused by moving objects in the scene has also been demonstrated.

REFERENCES

[1] J. Oliensis, "A multi-frame structure-from-motion algo-

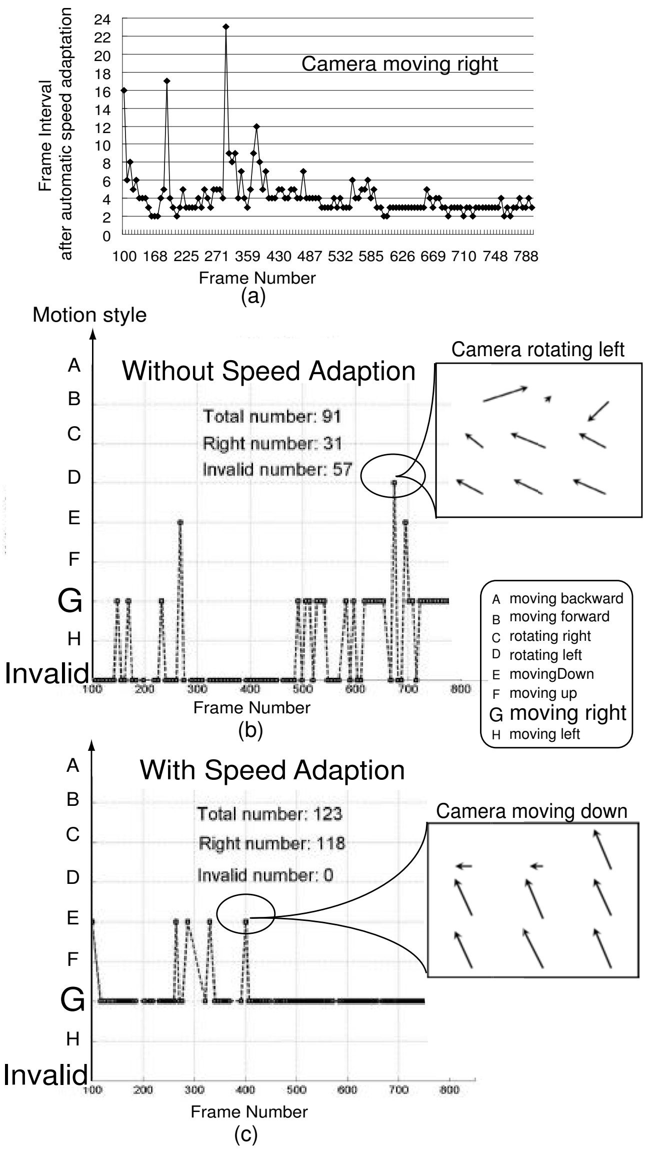


Figure 7: Results of introducing the speed-adaptive scheme to image sequence shown in Fig. 5(a).

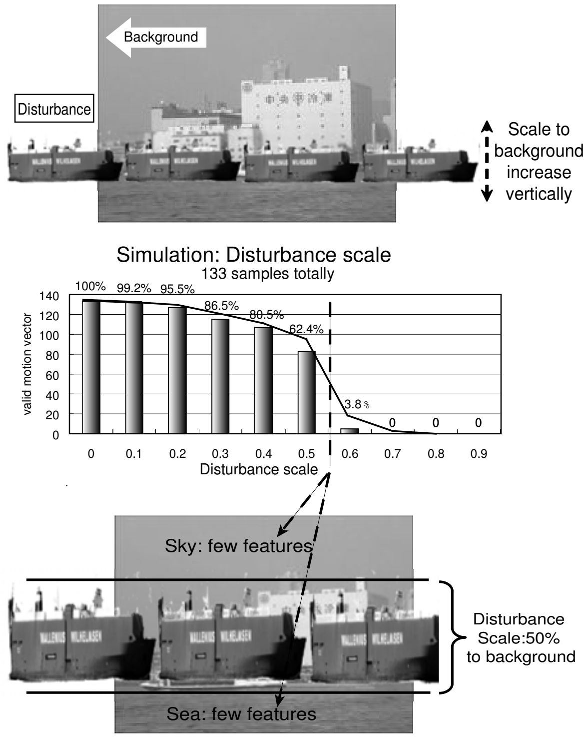


Figure 8: Anti-disturbance performance of the system. Here the image of ships with varying sizes was overlaid as an obstacle.

rithm under perspective projection," *Int. Journal Comp. Vision*, vol. 34, no. 2, pp. 163-192, 1999.

[2] Z. W. Kim and T. E. Cohn, "Pseudoreal-Time Activity Detection for Railroad Grade-Crossing Safety," *IEEE Trans. Intelligent Transportation*, vol. 5, no. 4, pp. 319-324, Dec. 2004.

[3] K. Daniilidis and H. H. Nagel, "Analytical results on error sensitivity of motion estimation from two views," *Image Vision Comput.*, vol. 8, pp. 297-303, 1990.

[4] D. Lelescu and D. Schonfeld, "Statistical sequential analysis for real-time video scene change detection on compressed multimedia bitstream," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 106-117, Mar. 2003.

[5] R. Vidal and S. Sastry, "Optimal segmentation of dynamic scenes from two perspective views," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 281-286.

[6] D. Comaniciu, V. Ramesh and P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, Jun. 2000, pp. 142-149.

[7] J. Hao and T. Shibata, "A VLSI-Implementation-Friendly Ego-Motion Detection Algorithm Based on Edge-Histogram Matching," in *Proc. IEEE ICASSP 2006*, Toulouse, France, May 14-18. 2006, pp. II-245-248.

[8] M. Yagi and T. Shibata, "An Image Representation Algorithm Compatible With Neural-Associative-Processor-Based Hardware Recognition Systems," *IEEE Trans. Neural Networks*, vol. 14, no. 5, pp. 1144-1161, Sep. 2003.

[9] Y. Suzuki and T. Shibata, "An edge-based face detection algorithm robust against illumination, focus, and scale variations," in *Proc. EUSIPCO 2004*, vol. 3, Vienna, Austria, September 6-10. 2004, pp. 2279-2282.

[10] H. Yamasaki and T. Shibata, "A Real-Time Image-Feature- Extraction and Vector-Generation VLSI Employing Arrayed- Shift-Register Architecture," in *Proc. ESSCIRC 2005*, Grenoble, France, September 12-16. 2005, pp. 121-124.