Associative Memory-Based Systems with Recognition and Learning Capability

Hans Juergen Mattausch, Tetsushi Koide
Research Center for Nanodeivices and Systems, Graduate School of Advanced Sciences of Matter, Hiroshima University
1-4-2 Kagamiyama, Higashi-Hiroshima 739-8527, Japan
Phone: +81-82-424-6265, FAX: +81-82-424-3499, e-mail: {hjm, koide}@sxsys.hiroshima-u.ac.jp

1. Introduction

Pattern recognition and learning are basic functions, which are needed to build artificial systems with capabilities similar to the human brain [1]. Their effective implementation in integrated circuits is therefore of great technical importance. We are developing a flexible memory-based architecture for this purpose, which can be expected to allow intelligent data processing similar to the human brain such as object-feature extraction, object recognition and learning or even judgment. In the COE program, we are mainly investigating the associative memory-based systems with recognition and learning capability (Fig. 1).

2. Associative Memory for Pattern Recognition

An important basic component for intelligent data processing is an associative memory with nearest-match capability between input-data words and a stored basis of reference-data words. Especially for real-time recognition it will be necessary to implement fast matching up to large absolute minimum distances. In this project, we have developed the fast and compact architecture for associative memories to efficiently implement the knowledge base of the system with learning and recognition capability. Key innovation is a fully-parallel, combined digital/analog realization of the search function, which allows short nearest-match times with the Hamming as well as Manhattan distance measures (Fig. 2) [3, 4]. The chosen associative-memory approach has in particular a high probability of being superior to the neural network approach, because there is no restriction on the type of the stored patterns. Therefore, a large variety of applications can be covered with the same hardware, which opens also the chance of adaptability (even self-adaptability) of the system to different applications. Furthermore, integration in conventional CMOS-technology is expected to be easy.

We have designed test chips in 0.6um (Hamming) [3] and in 0.35um (Manhattan) [9] CMOS technologies. The 9.75mm^2 Hamming test-chip with 32 reference patterns and 768 equivalent bit per pattern, has a performance of <70nsec nearest-match time, equivalent to a 32bit computer with 150GOPS/mm^2, at a power dissipation of 43mW. The 8.6mm^2 Manhattan test-chip with 128 reference patterns and 496 equivalent bit per pattern, has a performance of <190nsec nearest match time, equivalent to a 32bit computer with 20GOPS/mm^2, at a power dissipation of 91mW. We have also proposed a bank-type associative memory verified by test chips in 0.35um CMOS technology (Table 1, Fig. 3) [2]. This bank-type architecture extends the possibility of fully-parallel nearest-match search to an in principle infinite space of reference patterns. These obtained data are sufficient for application in high-performance mobile real-time systems such as systems for image compression by vector-quantization.

3. Automatic Associative Memory based Reference-Pattern Learning and Optimization

In this project, a system concept with realizes high-speed pattern matching and automatic pattern learning has been developed on the basis of an associative memory with short-term and long-term storage regions (Fig. 4) [8]. The applied learning algorithm (Fig. 5) uses a 4-step process for each learning cycle: (1) Nearest-match determination (winner) in the associative memory for an input pattern. (2) Decision whether the input pattern is known by the system on the basis of the winner distance. (3) Increasing memorization strength (rank in the storage space) of the winner if the input pattern is known. (4) Learning of the input pattern with a specific rank in the short-term memory and forgetting the reference pattern with the lowest rank in the short-term memory if the input pattern is not known (Fig. 4). An LSI architecture is proposed, which implements the developed algorithm with a fully-parallel associative memory, a rank-processing circuit and a control circuitry including digital winner-distance calculation (Fig. 6). A CMOS test-chip, which verifies the developed architecture for an associative-memory storage capacity of 64 pattern and Manhattan-distance search has been design and fabricated.

4. Low Power Real-time Image Segmentation

Image segmentation is the extraction process of all objects from natural input images and is the necessary first step of object-oriented image processing such as object recognition or object tracking. In this project, we have proposed a cell-network-based digital image segmentation algorithm/architecture with pixel parallel processing for gray-scale/color images in real-time applications (Fig. 7) [5, 6]. A CMOS test-chip for the cell-network, which is the main functional stage, has been fabricated, in a 0.35um CMOS technology and verifies the effectiveness of our proposal. In the performance verification of the test-chip, high speed segmentation in <9.5usec and low power dissipation of <36.4mW@10MHz are measured. The extrapolation results to larger image sizes suggest, that QVGA-size image segmentation will be possible within 300usec @10MHz at the 90nm CMOS technology node. Furthermore, we have proposed a low-power and hardware-efficient pipelined segmentation architecture for VGA-size motion pictures, which applies a subdivided-image approach (SIA) for compact implementation and a boundary-active-only (BAO) scheme for low-power dissipation [7]. We have verified the effectiveness of the proposed architecture with a 51mm^2 test-circuit in 0.35um CMOS technology for the segmentation-network core consisting of 41x33 cells (Fig. 8). The segmentation performance for a VGA-size input image is 21.8mW power dissipation and 7.49msec segmentation time at 10MHz clock frequency.

5. Conclusion

We have introduced our reseaches for an associative memory-based system with recognition and learning capability. The next steps in our research effort towards the complete system include architecture/circuit development for adaptive pattern learning unit and feature-extraction unit, which requires also the selection of concreate application examples, and an prototype development of a system with recognition and learning capability.

Acknowledgments: The test-chips in this study have been fabricated in the chip fabrication program of VLSI Design Education Center (VDEC), the University of Tokyo in the collaboration with Rohm Corporation and Toppan Printing Corporation.

References

[1] D. R. Tveter, The Pattern Recognition Basis of Artificial Intelligence, Los Alamitos, CA: IEEE Computer Society, 1998.
[2] T. Koide, et al., Ext. Abst. SSDM2004, pp. 360-361, 2004.
[3] H. J. Mattausch et al., IEEE JSSC, vol. 37, 2, pp.218-227, 2002.
[4] H. J. Mattausch et al., Symp. on VLSI Circuits, pp. 252-255, 2002.
[5] T. Morimoto, et al., Ext. Abst. SSDM2002, pp. 242-243, 2002.
[6] T. Morimoto, et al., IEICE Trans. E87-D, 2, pp. 500-503, 2004.
[7] T. Morimoto, et al., Ext. Abst. SSDM2004, pp. 138-139, 2004.
[8] Y. Shirakawa, et al., Ext. Abst. SSDM2004, pp. 362-363, 2004.
[9] Y. Yano, et al., Ext. Abst. SSDM2002, pp. 254-255, 2002.

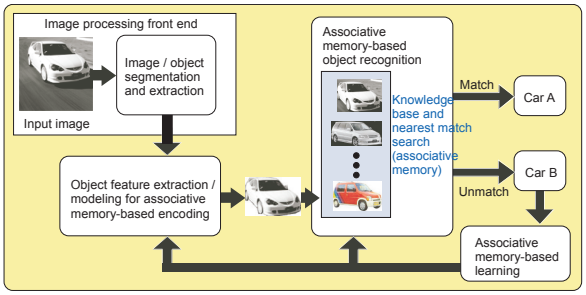


Figure 1: Structure of envisaged associative memory-based systems for the case of a visual input and illustrated with the example of recognizing and learning different types of cars.

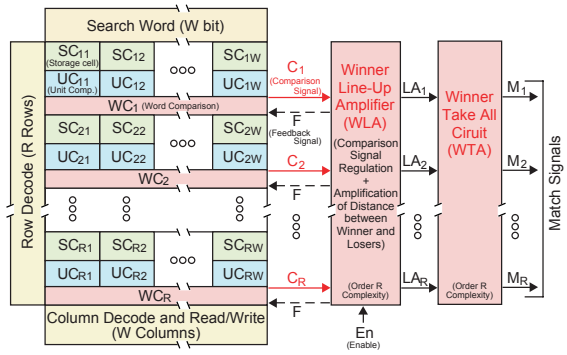


Figure 2: Block diagram of the compact-associative-memory architecture with fast fully-parallel match capability according to the Hamming/Manhattan distance.

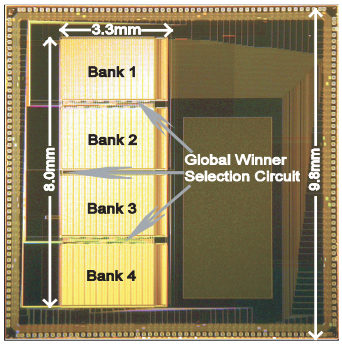


Figure 3: Chip photo of the 4-bank associative-memory (0.35um CMOS).

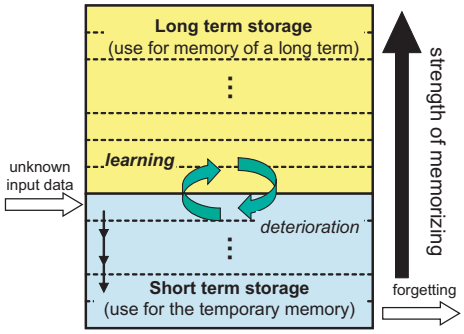


Figure 4: Learning concept based on a short/long term memory.

Table 1: Characteristics of the bank-type Manhattan-distance associative memory.

Distance Measure	Manhattan (5 bit)	
	2-Bank	4-Bank
Reference Number	128 (64 x 2)	256 (64 x 4)
Design Area	11.8mm^2	26.5mm^2
Search Unit Area	0.99mm^2	1.97mm^2
Search Range	0 - 496bit	0 - 496bit
Winner-Search Time (Simulation)	< 260nsec	< 280nsec
Power Dissipation (Simulation)	< 330mW	< 640mW
Performance	128 GOPS	229 GOPS
Technology	0.35µm 3-metal CMOS	0.35µm 3-metal CMOS
Supply Voltage	3.3V	3.3V

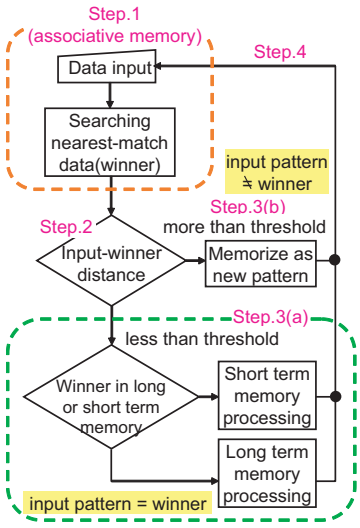


Figure 5: Flow chart of proposed associative-memory-based pattern-learning algorithm.

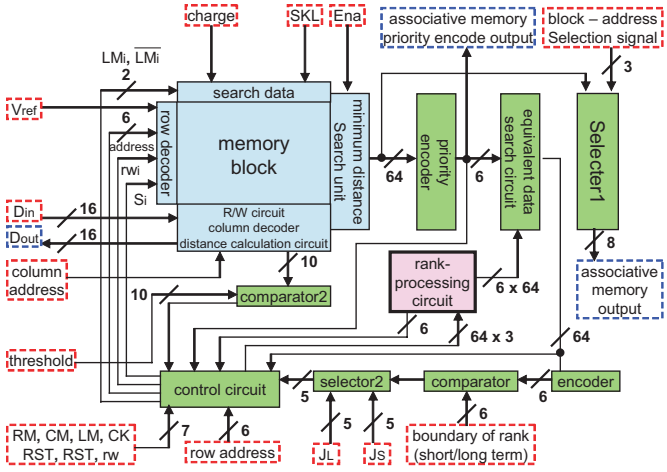


Figure 6: Associative-memory-based automatic pattern learning architecture with 64 patterns. Long/short-term-memory size, parameters J_L, J_s and the threshold in the algorithm can be set externally.

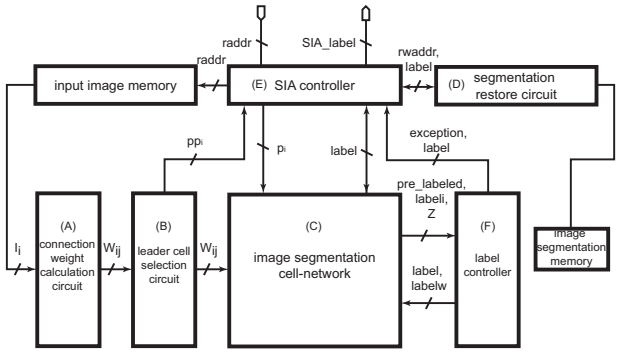


Figure 7: Block diagram of the cell-network-based image segmentation architecture with subdivided-image approach (SIA).

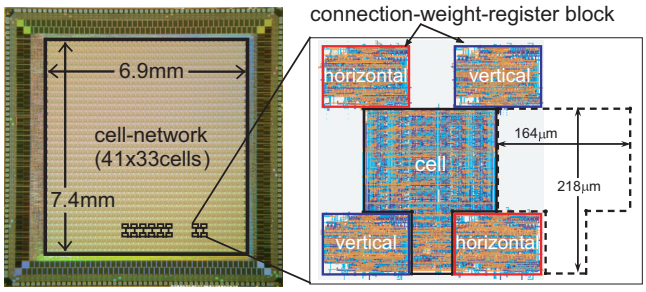
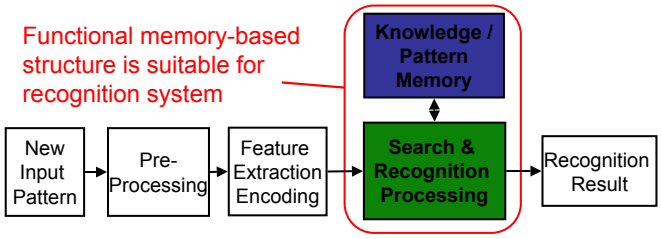


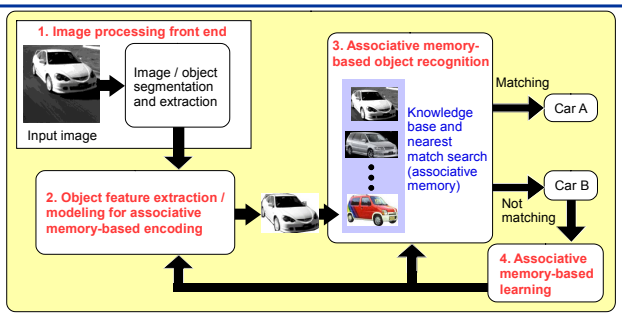
Figure 8: Die photo of the cell-network with BAO including 41x33 cells designed in a 0.35um 3-metal CMOS technology. The layout of cell and connection-weight-register blocks is magnified on the right side.

Research Objective

- Design of associative-memory-based system with reference pattern learning and recognition
- Development of an associative-memory and picture-segmentation-based tracking architecture for moving objects
- Investigation of realization with 3DCSS platform



Envisaged System Architecture



1. Extract the object of interest from the input data. For an image as input data, this stage requires an image segmentation function and a procedure for selecting the segment (or object) of interest.
2. Prepares the data of the selected object for a comparison with the knowledge base of the system by extracting the objects characteristic features.
3. Knowledge base of the system which includes a search function for finding the best match to an input pattern from the 2nd stage.
4. The learning stage includes a feedback to the 3rd stage, the knowledge base, and possibly also to the 2nd stage for the characteristic-feature extraction.

Architecture Realization by 3D Custom Stack System (3DCSS)

- Realization of Tbit processing (~1Tbit/sec) with a large capacity bank-type associative memory
- Pipeline processing over multiple chips
- Development of an integration architecture for a bank-type associative memory with learning and recognition capability
- Improvement of the automatic reference pattern learning and optimization algorithms
- Application to a moving object tracking architecture based on associative memory and picture segmentation

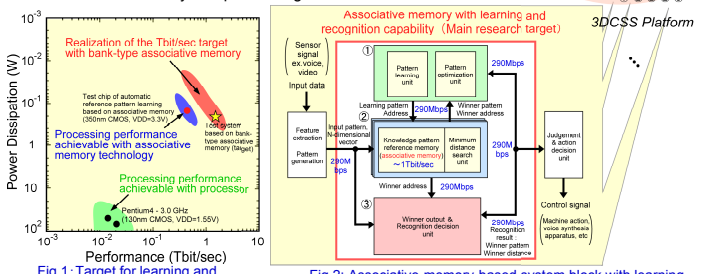


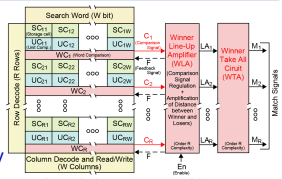
Fig.1 : Target for learning and recognition processing performance.

Fig.2: Associative-memory-based system block with learning and recognition capability chosen as integration target

Associative Memory for Pattern Recognition

○ Research Contents

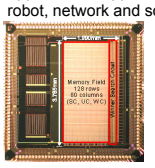
- Fast search capability with a mixed digital/analog circuit
- Distance amplification circuit with self-adaptable operating point depending on input conditions
- Distance measure encoding scheme for adaptability to applications



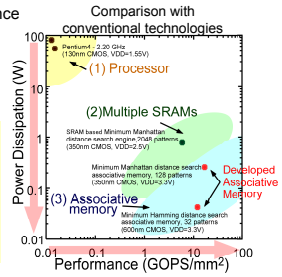
Associative memory architecture

○ Special Features of Developed Associative Memory

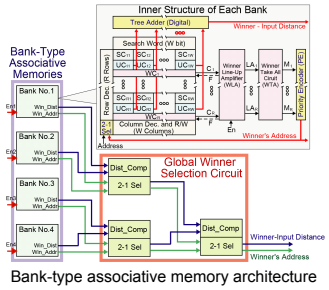
- High speed, fully-parallel minimum-distance Hamming / Manhattan distance search
- High area efficiency and low power dissipation per reference pattern
- High reliability and design in conventional CMOS technology
- Applicable to applications such as artificial intelligence, robot, network and so on.



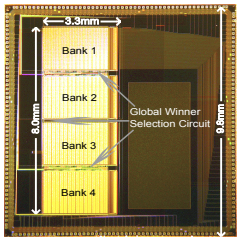
- Minimum-distance search time: < 240ns
- Power dissipation : <260mW at 10MHz (34.7mW/mm²)
- Performance : 170GOPS (20GOPS/mm²)
- Search time in practical applications such as image compression: < 150ns



Bank-Type Fully-Parallel Associative Memory



Bank-type associative memory architecture



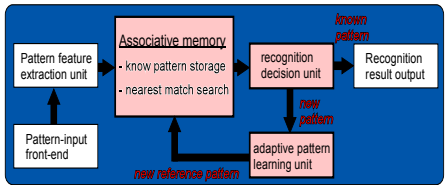
Test chip of the 4-bank associative memory (0.35um CMOS)

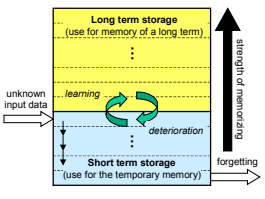
Manhattan(5bit)	Ref. Number	Design Area (mm^2)	Search Unit Area (mm^2)	Search Range (bit)	Winner-Search Time (simulation)	Power Dissipation (simulation)	Performance	Technology	Supply Voltage
2-bank	128 (64x2)	11.8	0.99	0 - 496	< 260nsec	< 330mW	128GOPS	0.35umCMOS	3.3V
4-bank	256 (64x4)	26.5	1.97	0 - 496	< 280nsec	< 640mW	229GOPS	0.35umCMOS	3.3V

The proposed architecture extends the possibility of fully-parallel nearest-match search to an in principle infinite space of reference patterns. For search problems with categorizable reference-data space the power dissipation can be reduced to the value for one bank in the best case.

Associative Memory based System with Learning Capability

- New Learning Method, surpassing the capabilities of neural networks.
- Learning algorithm and architecture which implement the short-term and long-term memory principle of the human brain.
- Architecture which is suitable for LSI integration.





Learning unit of suitable patterns

(1) Distance information (Hamming/Manhattan Distance)
(2) Scheme with 1 or more thresholds
(3) Algorithm based on similarity to known data
(4) Extrapolation from previous to future input data

■Expressing the memorization strength of the associative memory's reference data with a rank.
■The upper (yellow region) and lower (blue region) ranks model the long- and short-term storage, respectively.

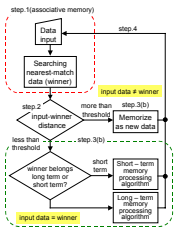
Automatic Associative Learning of Reference Pattern

○ Research Contents

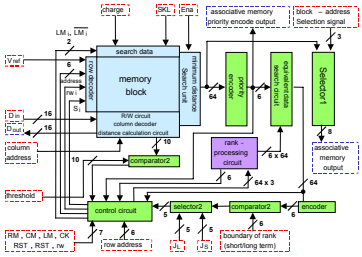
・Development of an LSI-architecture capable of pattern matching and pattern learning without teacher

○ Results

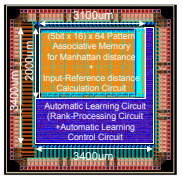
・Algorithm based on associative memory
・Realization of short- and long-term memory function
・LSI architecture and test-chip in 0.35um CMOS



(a) 4-step algorithm for realization of pattern learning



(b) LSI architecture for implementation of the learning algorithm



- Area : 2.0mmx3.4mm (64 pattern capacity)
- Learning cycle : 300ns
- Programmable short- and long-term storage size
- Associative memory with Manhattan-distance measure
- Data processing in 1-learning cycle : ～0.11Tbit/sec

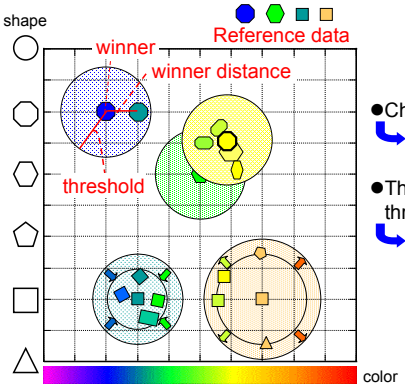
(c) LSI chip realizing automatic data-base learning (0.35um 3-metal CMOS)

Reference-Pattern Learning and Optimization for Associative-Memory-Based Pattern-Recognition Systems

Improvement in the rate of recognition of an associative memory → Learning

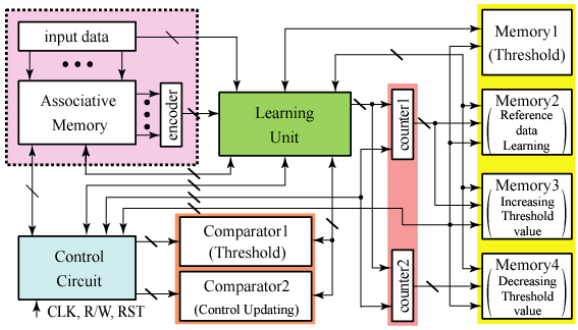
Reference data most similar to input data = winner

The distance between input and winner = winner distance



- Change input data distribution → Optimization of reference pattern
- Threshold is too large for a proper threshold or too small → Optimization of threshold

Chip Architecture for VLSI Implementation of Reference-Pattern Learning and Optimization Algorithm



- Pattern matching with fully parallel associative memory
- Distance measure is the Manhattan distance
- Complex hardware (multiplier) is not necessary

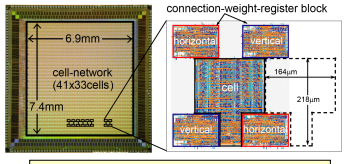
Low Power Real-time Image Segmentation

○ Research Contents

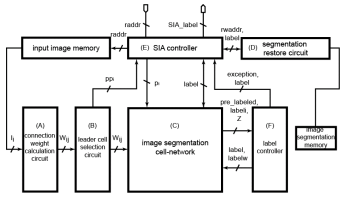
・Development of low power real-time color image segmentation chip architecture

○ Results

・Digital color image segmentation algorithm based on region growing approach
・Fully pixel parallel architecture based on cell network
・Low power operation with boundary-active-only scheme
・Proposal of architecture with pipeline processing of tiled images for large-scale image segmentation

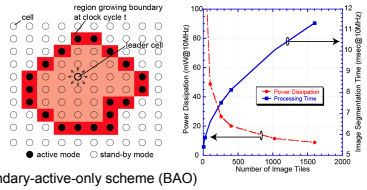


- Image segmentation test chip (0.35um 3-metal CMOS)
- Area : 6.9mmx7.4mm (41x33cells)
- Segmentation time : 23usec@10MHz (41x33pixels)
- Power dissipation : < 30mW@10MHz



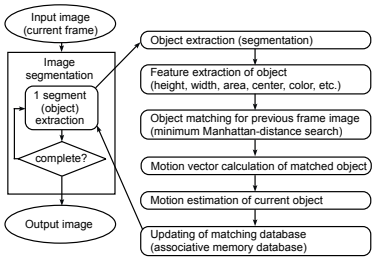
Architecture block diagram

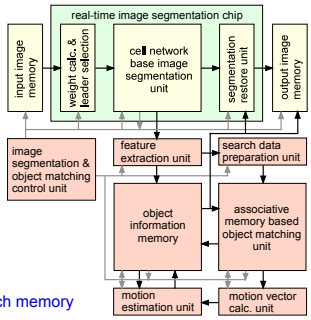
Estimated performance for VGA images as a function of tile size



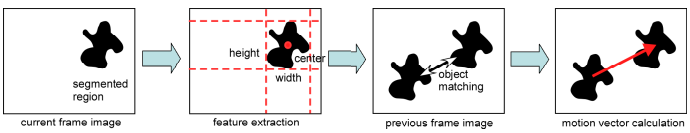
Boundary-active-only scheme (BAO)

Object Tracking using Image Segmentation and Pattern Matching





✓ Object extraction with image segmentation
✓ Object matching based on minimum-distance search memory



Conclusion

- Associative Memory for Pattern Recognition
  - Fully parallel mixed analog-digital architecture developed and verified
  - Arbitrarily large reference pattern space realizable with bank-type architecture
- Low Power Real-time Image Segmentation
  - Low-power region-growing algorithm with real-time capability developed and tested
  - VLSI integration for large image sizes with conventional CMOS technology verified by test chip design
- Automatic Associative Memory based Reference-Pattern Learning and Optimization
  - Algorithms for continuous pattern learning and pattern optimization developed and tested
  - VLSI integration suitability verified by test chip design
- Object Tracking using Image Segmentation and Pattern Matching
  - Algorithm capable of multi-object tracking, even for the moving camera case, developed and tested
  - VLSI architecture for tracking-system realization developed and under verification