Associative Memory-Based Systems With Recognition and Learning Capability

Tetsushi Koide and Hans Jürgen Mattausch

System Design and Architecture Research Division, Research Center for Nanodevices and Systems, Hiroshima University
1-4-2 Kagamiyama, Higashi-Hiroshima 739-8527, Japan
Phone: +81-824-24-6265 Fax: +81-824-22-7185 e-mail: {koide, hjm}@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. The architecture is envisaged to improve the traditional neural network approach, which emphasizes the role of the processing elements (neurons) and their interconnection network, but neglects sufficient exploitation of the third powerful system-component, which is the memory.

2. Envisaged System Structure

The aimed-at system structure is depicted in Fig. 1 for the case of a visual input and is illustrated with the example of recognizing and learning different types of cars. It contains 4 functional stages, for which efficient solutions have to be found. The front-end stage has the task to extract the object of interest from the input data [2]. For an image as input data, this stage requires an image segmentation function and a procedure for selecting the segment (or object) of interest, as will be explained in more detail in section 4. The second stage prepares the data of the selected object for a comparison with the knowledge base of the system by extracting the objects characteristic features. The detailed construction of this second stage is extremely dependent on the type of object and knowledge base involved in the recognition/learning process. Ideally a solution, which can adapt itself to different types of objects, is desirable. The third stage represents the knowledge base of the system and must include a search function for finding the best match, known by the system, to an input pattern from the second stage. Especially the best-match-search function requires an enormous computational effort and an efficient solution is a crucial issue. We discuss our associative-memory solution to this problem in more detail in section 3. The final learning stage includes a feedback to the third stage, the knowledge base, and possibly also to the second stage for the characteristic-feature extraction. It must decide two issues: the addition of new objects two the knowledge base as well as possible modifications of the characteristic-feature extraction.

3. Associative Memory Core

We have developed the fast and compact architecture for associative memories of Fig. 2, to efficiently implement the knowledge base of the envisaged system with learning and recognition capability [3, 4]. Key innovation is a fully-parallel, combined digital/analog realization of the search function, which allows short nearest-match times (<100ns) with the Hamming as well as Manhattan distance measures. 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.

Figures 3 and 4 show structure and test-chip implementation of a pattern-matching engine with self-adaptability to Hamming or Manhattan distance, which exploits our new associative-memory architecture. The self-adaptability is realized by distance-measure encoding as shown in Fig. 3. The test-chip of Fig. 4 achieves very high performance data of $> 10^7$ pattern/sec throughput, corresponding to about 1 TOPS processing power, at very low power dissipation of 43 mW [4]. A direct realization of the Manhattan distance by implementing the subtraction and absolute value circuitry into the unit-comparison part (UC in Fig. 2) of the associative-memory architecture is also possible (see Refs. [5, 6]). The application examples of Japanese character recognition for a Hamming-distance associative memory and image compression by codebook-based vector quantization [7] are depicted in Fig. 5(a) and 5(b), respectively.

4. Image-Preprocessing Front End

The structure of a possible front-end for image processing is explained in more detail as part of the moving object and tracking/recognition system for intelligent vehicle guidance in Fig. 6. Such an intelligent vehicle-guidance system can be regarded as a special application of the general system architecture with recognition and learning capability, shown in Fig. 1. Core of the image-processing front end is the image segmentation and extraction block. We have developed a digital CMOS-based architecture for real-time color-motion-picture segmentation, which is explained in more detail in Refs. [8, 9, 10].

5. Implementation of the Learning Function

Our planned architecture for implementing the learning function is depicted in Fig. 7. It utilizes the associative-memory core explained in section 3, a recognition-decision unit and an adaptive pattern-learning unit in a feedback-loop connection. For each input pattern from the feature-extraction unit, the associative-memory core will search for the pattern with minimum distance within its knowledge base.

The nearest-match pattern and its corresponding distance to the input pattern will then be transferred to the recognition-decision unit. On the basis of the distance information and a single- or multiple-threshold scheme, the recognition-decision unit will decide whether the input pattern is a known pattern or an unknown pattern. If the input pattern is found to be a known pattern, the recognition result, namely the information associated with the input pattern and possibly an error probability information, will be the system output. On the other hand, if the input pattern is found to be unknown, the adaptive pattern-learning unit will be activated.

The adaptive pattern-learning unit will use the distance information, knowledge-base-related similarity-reasoning algorithms and information about unknown patterns, which appeared previously as inputs, and will decide whether the present unknown pattern should be added to the knowledge base of the associative memory. Once a positive decision is reached, the update of the associative-memory core is done with a simple writing operation. Presently, we have not yet finished the development of the detailed algorithms and the corresponding CMOS circuits for the adaptive pattern-learning unit. This will therefore be one of our challenging research subjects in the near future.

6. Conclusion

We have presented our architecture concepts for an associate-memory-based system with recognition and learning capability as well as our present development status. A number of important milestones for making these architecture concepts a reality have been reached. Most important research results are architecture development and CMOS implementation of:

- a compact associative-memory core with fully-parallel nearest match function [3, 4, 5, 6],

- a digital real-time color-motion-picture segmentation front-end [8, 9, 10, 11].

The next steps in our research effort towards the complete system include:

- architecture/circuit development for the adaptive pattern-learning unit
- architecture/circuit development for the feature-extraction unit, which requires also the selection of concrete application examples
- prototype development of a complete system with recognition and learning capability.

References

[1] D. R. Tveter, The Pattern Recognition Basis of Artificial Intelligence, Los Alamitos, CA: IEEE Computer Society, 1998.
[2] J. C. Russ, The Image Processing Handbook, CRC PRESS, 1999.
[3] H.J. Mattausch et al., IEEE J. Solid-State Circuits, vol. 37, pp.218-227, 2002.
[4] H.J. Mattausch et al., Symposium on VLSI Circuits, pp. 252-255, 2002.
[5] Y. Yano, et al., Extend. Abst. of the 2002 international Conf. on Solid State Devices and Materials, pp. 254-255, 2002.
[6] Y. Yano, et al., Proc. of 1st Hiroshima Int'l Workshop on Nanoelectronics for Terra-Bit Information Processing, 2003.
[7] A. Gersho, et al., Vector Quantization and Signal Compression, Boston, MA: Kluwer Academic, 1992.
[8] T. Morimoto, et al., Extend. Abst. of the 2002 international Conf. on Solid State Devices and Materials, pp. 242-243, 2002.
[9] T. Morimoto, et al., Proc. of 2002 IEEE Asia-Pacific Conf. on ASIC, pp. 237-240, 2002.
[10] T. Morimoto, et al., Proc. of 1st Hiroshima Int'l Workshop on Nanoelectronics for Terra-Bit Information Processing, 2003.
[11] T. Koide, et al., Proc. of Int'l Tech. Conf. on Circuits/Systs., Computers and Communications, pp. 670-673, 2002.

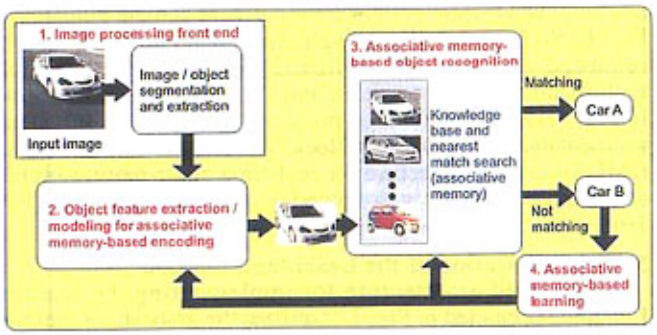


Fig. 1: Structure of envisaged associative memory-based systems with recognition and learning capability.

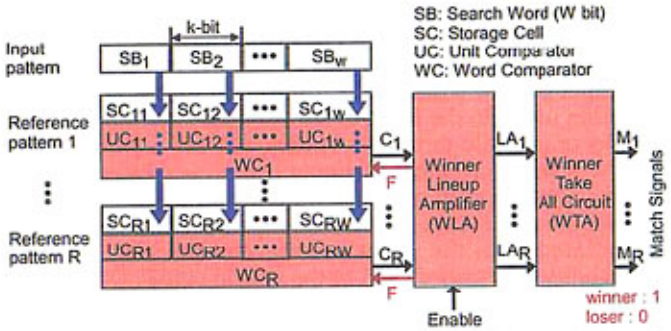


Fig. 2 : Associative memory architecture with fast fully-parallel match capability. Reference pattern storage and nearest match search circuits are integrated.

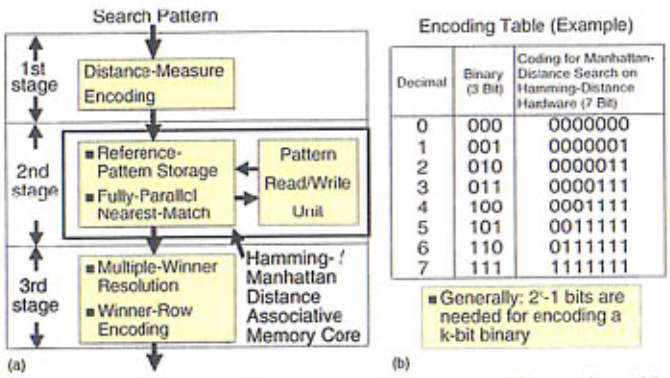


Fig. 3 : Block diagram of the proposed pattern-matching engine with adaptable distance measure. Distance-measure encoding allows to use the fastest distance measure in the core and to handle the additional complexity of other distance measures in the encoding stage.

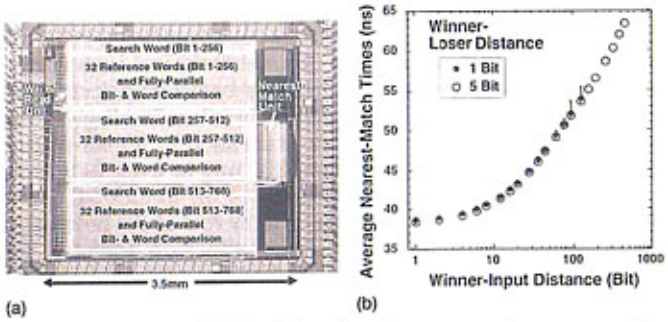
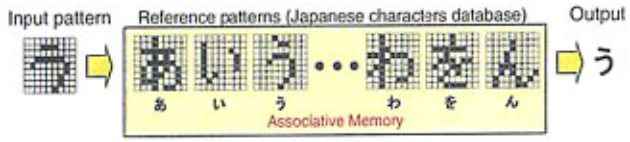
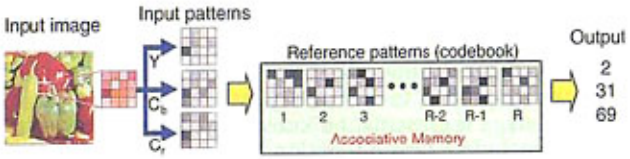


Fig. 4 : 32-pattern, 24K-bit Hamming-distance search memory chip. (a) Photomicrograph of test-chip fabricated in 0.6μm CMOS technology. (b) Nearest-search times are < 70ns and power dissipation is 43mW at 10MHz (4.4mW/mm^2).



(a) Image pattern recognition (Hamming-Distance).



(b) Codebook-based data compression (Manhattan-Distance).

Fig. 5: Associative memory-based application examples. (a) for Hamming-disrtance search, (b) for Manhattan-distance search.

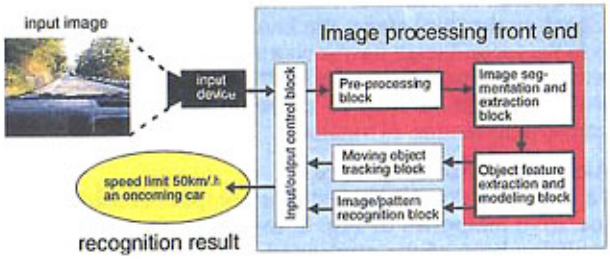


Fig. 6: Image segmentation and feature extraction for real time applications.

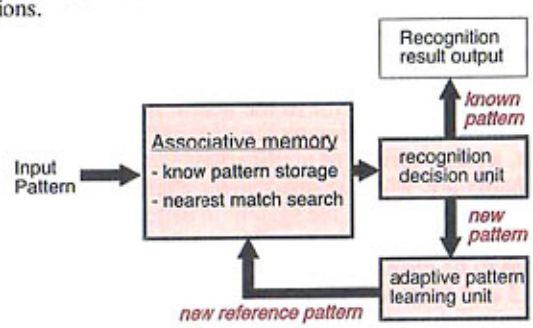


Fig. 7: Associative memory-based adaptive learning. Minimum distance search functionality enables learning capability in the associative-memory-based systems.

Associative Memory-Based Systems With Recognition And Learning Capability

Tetsushi Koide and Hans Juergen Mattaush

Research Center for Nanodevices and Systems (RCNS)
Hiroshima University

http://www.rcns.hiroshima-u.ac.jp/

1. Introduction

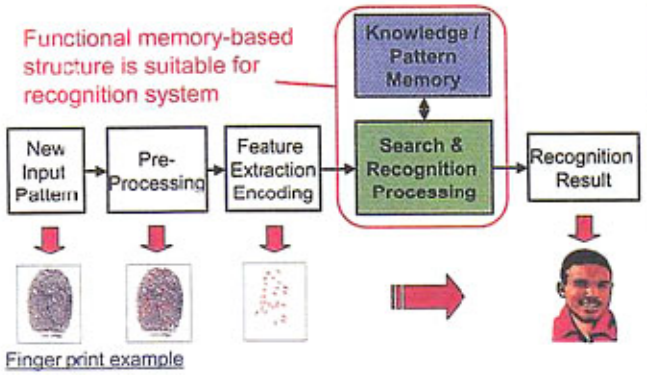
◆ Pattern recognition and learning are basic functions, which are needed to build artificial systems with capabilities similar to the human brain. Their effective implementation in integrated circuits is therefore of great technical importance.

◆ We are developing a flexible memory-based architecture for this purpose, which can be expected to allow intelligent processing (object-feature extraction / recognition, learning, judgment)

◆ Traditional neural network approach emphasizes the role of the processing elements (neurons) and their interconnection network, but neglects sufficient exploitation of the third powerful system-component, which is the memory

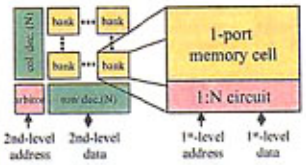
Functional Memory Based Intelligent Information System

2. Basic Structure Of Recognition System



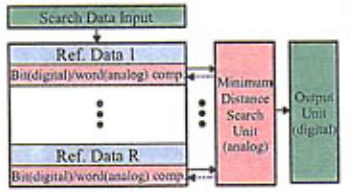
3. Developed Functional Memories

Multi-port Memory
High-speed parallel access



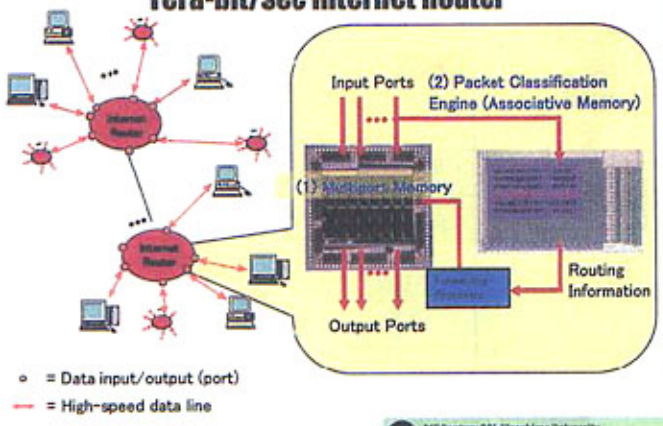
- High area efficiency
- High scalability (port/memory size)
- High input/output band-width
- Short latency

Associative Memory
Fully-parallel pattern matching

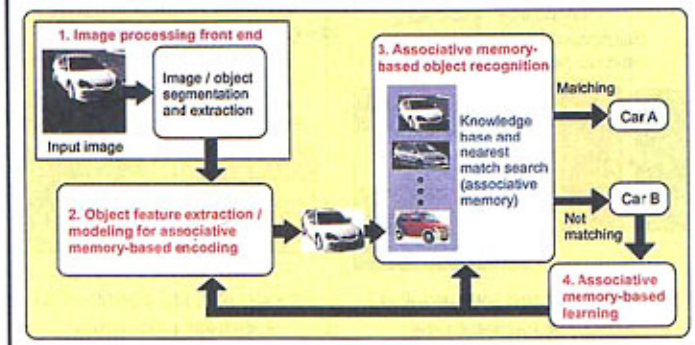


- Analog/digital mixed circuits
- High-speed pattern matching
- Low power dissipation
- Compact
- High-reliability (feedback control)

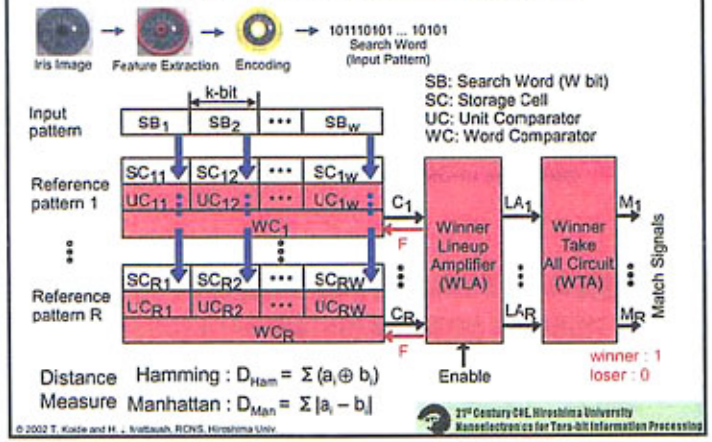
4. Example of Functional Memory-Base Application : Tera-bit/sec Internet Router



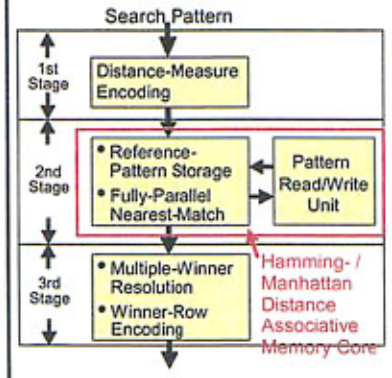
5. Structure Of Envisaged Associative Memory-Based Systems With Recognition And Learning Capability



6. Associative Memory Architecture With Fast Fully-Parallel Match Capability



7. Structure Of Pattern Matching Engine Based On Associative Memory Core

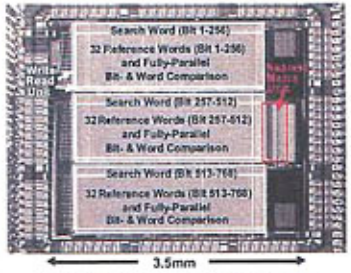


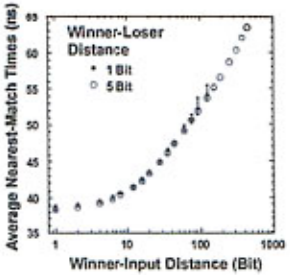
Encoding Table (Example)

Decimal	Binary (3 Bit)	Coding for Manhattan-Distance Search on Hamming-Distance Hardware (7 Bit)
0	000	0000000
1	001	0000001
2	010	0000011
3	011	0000111
4	100	0001111
5	101	0011111
6	110	0111111
7	111	1111111

• Generally: 2^k-1 bits are needed for encoding a k-bit binary

8. 32-pattern, 24K-bit Hamming-Distance Search Memory Test Chip

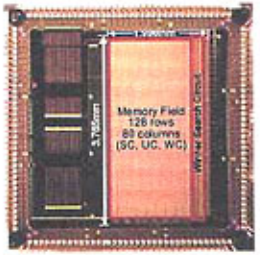


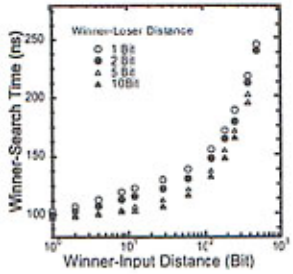


Test-chip fabricated in 0.6μm CMOS

- Nearest-search times : < 70ns
- Power dissipation : 43mW at 10MHz (4.4mW/mm²)

9. 128-Pattern, 10K-Bit Manhattan-Distance Search Memory Test Chip



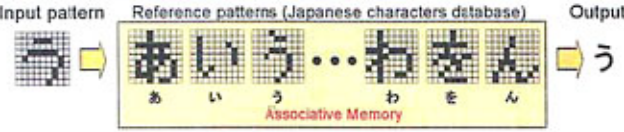


Test-chip fabricated in 0.35μm CMOS

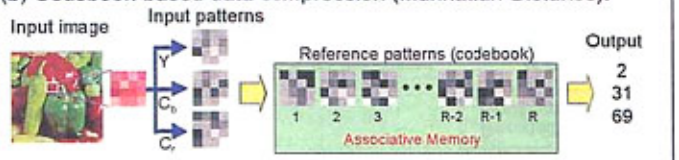
- Nearest-search times : < 240ns
- Power cissipation : <260mW at 10MHz (34.7mW/mm²)
- Equivalent performance : 170GOPS (20GOPS/mm²)

10. Associative Memory-Based Application Examples

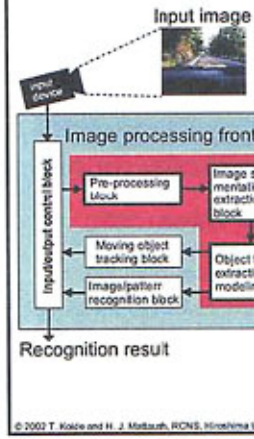
(a) Image pattern recognition (Hamming-Distance).



(b) Codebook-based data compression (Manhattan-Distance).



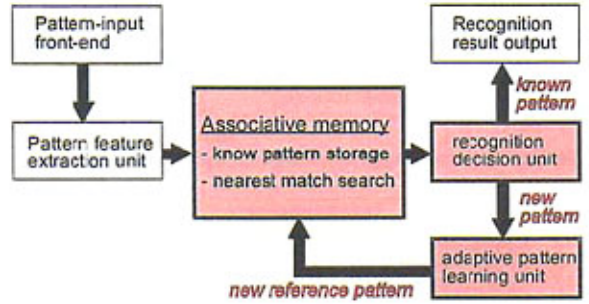
11. Image Segmentation And Feature Extraction For Real Time Applications



Core of the image-processing front end : image segmentation and extraction block

Developed test-chip of a digital CMOS-based architecture for real-time color-motion-picture segmentation

12. Associative Memory-Based Adaptive Learning



Adaptive pattern-learning unit will use the distance information, knowledge-base-related similarity-reasoning algorithms and information about unknown patterns

13. Conclusion

Most important research results are architecture development and CMOS implementation of :

1. Compact associative-memory core with fully-parallel nearest match function for Hamming & Manhattan distances
2. Digital real-time color-motion-picture segmentation front-end

The next steps effort towards the complete system include :

1. Architecture/circuit development for the adaptive pattern-learning unit
2. Architecture/circuit development for the feature-extraction unit, which requires the selection of concrete application examples
3. Prototype development of a complete system with recognition and learning capability