Systems with Recognition and Learning Capability Based on Associative Memory

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

The effective implementation of pattern recognition and learning, which are basic functions for building artificial systems with capabilities similar to the human brain [1], is of great technical and practical importance. For this purpose our research group is developing a flexible architecture, which is based on the minimum-distance-search associative memory as a core element. A major first task is to efficiently implement the additional capabilities of recognition, learning and also judgement into the associative memory. Furthermore, a sufficient variety of distance measures for the basic pattern-matching proces has to be realized. In the COE program, we are mainly investigating such an associative memory-based systems, as schematically depicted in Fig. 1, to enable intelligent data processing similar to the human brain such as object-feature extraction, object recognition and learning or even judgement.

2. Distance Measure of an Associative Memory for Efficient Pattern Recognition

An associative memory has the capability of determining the nearest match between input-data words and a stored basis of reference-data words according to a distance measure. 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 a fully-parallel, combined digital/analog realization of the associative memory's 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 and integration in conventional CMOS-technology is easy.

We have tested our architecture with chip designs in 0.6um (Hamming) [3] and in 0.35um (Manhattan) [9] CMOS technologies. A performance up to the equivalent of a 32bit computer with 150GOPS/mm^2 at low power dissipation of a few mW per mm^2 could be achieved. We have also proposed and verified a bank-type associative memory verified with test chips in 0.35um CMOS technology [2]. This bank-type architecture extends the possibility of fully-parallel nearest-match search to an in principle infinite space of reference patterns.

Recently, we have extended our associative-memory architecture to the realization of the Euclidean distance, which gives the correct distance between 2 points in vector space. Key points in our solution (Fig. 3) are the application of an analog squarer for each vector component and the avoidance of the square-root calculation, which has no influence on the winner determination.

3. Learning and Optimization of Reference-Patterns

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). Furthermore, an optimization architecture for the learned reference patterns is developed [10].

A CMOS test chip, which implements a fully-parallel associative memory with 64 patterns, the pattern-learning algorithm and the pattern-optimization algorithm has been designed and fabricated (Fig. 6) and is now under measurement.

4. Cell-Network Based 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

An overview over our reseach work on an associative memory-based system with recognition and learning capability has been given. The next steps towards the complete system include architecture/circuit development for the adaptive pattern learning unit and the feature-extraction unit. This requires also the selection of concrete application examples and the development of a prototype 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.
[10] Y. Shirakawa, et al., Proc. MWSCAS2004, pp. I561-I564, 2004.

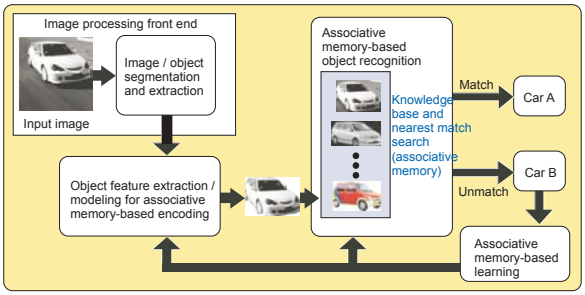


Figure 1: Structure of envisaged associative memory-based systems for the case of a visual input and illustrated with the example of recognizing and learning different types of cars.

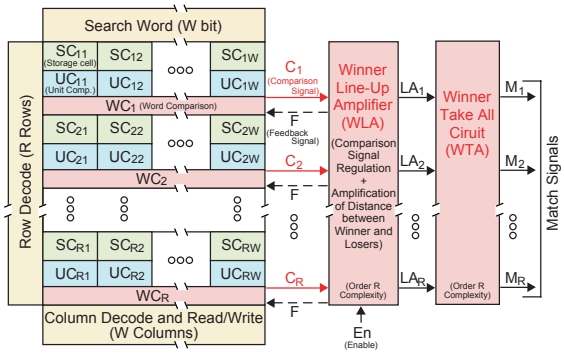


Figure 2: Block diagram of the compact-associative-memory architecture with fast fully-parallel match capability according to the Hamming/Manhattan distance.

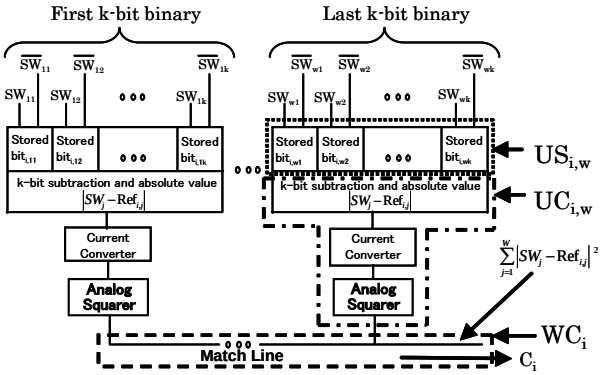


Figure 3: Architecture concept for an associative memory with fully-parallel minimum Euclidean-distance search.

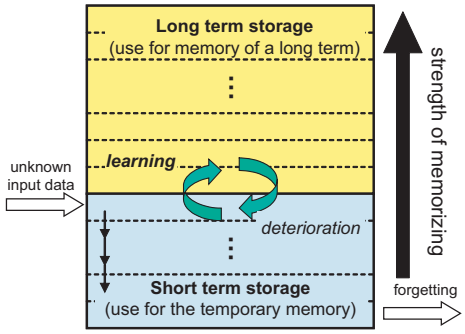


Figure 4: Learning concept based on a short/long term memory.

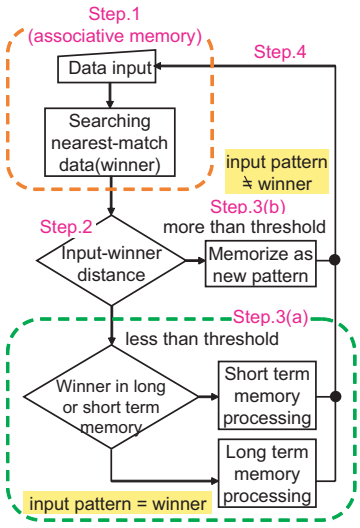


Figure 5: Flow chart of proposed associative-memory-based pattern-learning algorithm.

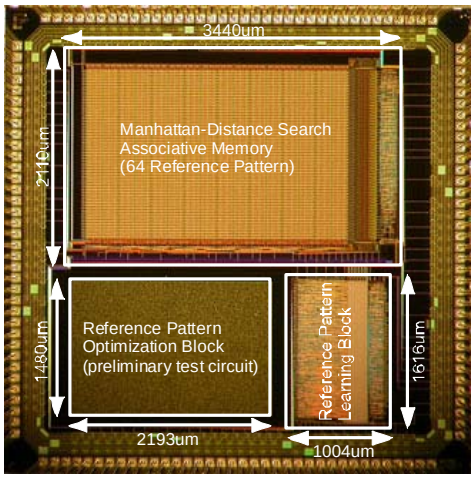


Figure 6: Associative-memory-based automatic pattern learning chip with 64 patterns. Long/short-term-memory size as well as the recognition threshold in the algorithm can be set externally.

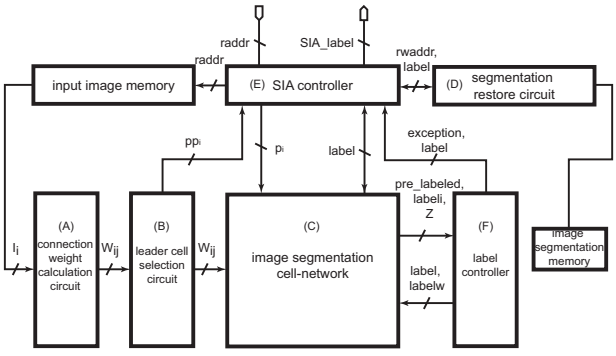


Figure 7: Block diagram of the cell-network-based image segmentation architecture with subdivided-image approach (SIA).

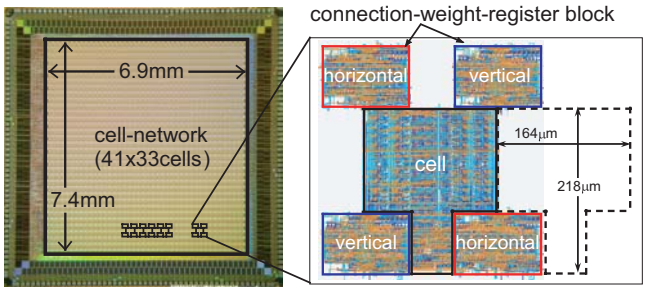


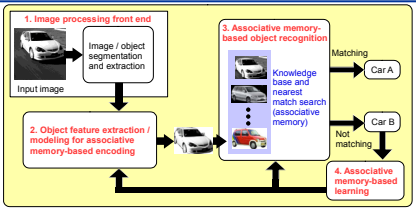
Figure 8: Die photo of the cell-network with BAO including 41x33 cells designed in a 0.35um 3-metal CMOS technology. The layout of cell and connection-weight-register blocks is magnified on the right side.

Systems with Recognition and Learning Capability Based on Associative Memory

Hans Jürgen Mattausch and Tetsushi Koide
Research Center for Nanodevices and Systems, Hiroshima University

Hiroshima University
Research Center for Nanodevices and Systems (RCNS)
System Design and Architecture Research Division

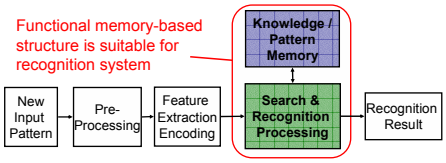
Envisaged System Architecture



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.

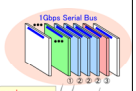
Research Objective

- 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



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



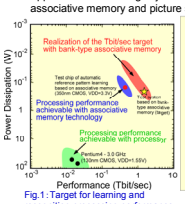


Fig.1: Target for learning and recognition processing performance.

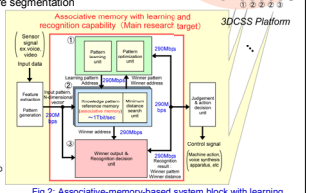
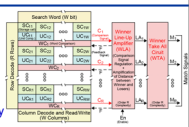


Fig.2: Associative-memory-based system block with learning and recognition capability chosen as integration target

Associative Memory for Pattern Recognition

Research Contents

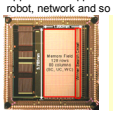
- Fast search capability with a mixed digital/analog circuit
- Distance amplification circuit with self-adaptable operating point depending on input conditions
- Distance measure encoding scheme for adaptability to applications



Associative memory architecture

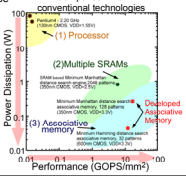
Special Features of Developed Associative Memory

- 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.

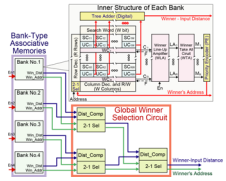


- Minimum-distance search time: < 240ns
- Power dissipation : <260mW at 10MHz (34.7mW/mm²)
- Performance : 170GOPS (200GOPS/mm²)
- Search time in practical applications such as image compression: < 150ns

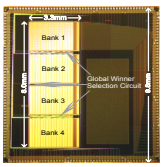
Comparison with conventional technologies



Bank-Type Fully-Parallel Associative Memory



Bank-type associative memory architecture

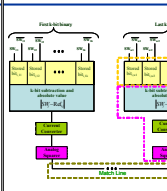


Test chip of the 4-bank associative memory (0.35um CMOS)

Manhattan(Sbit)	Ref. Number	Design Area (mm²)	Search Unit Area (mm²)	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

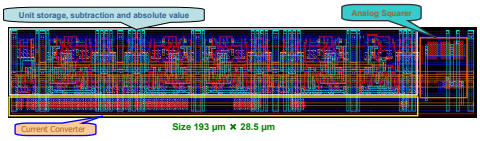
The proposed architecture extends the possibility of fully-parallel nearest-match search to an in principle infinite space of reference patterns. For search problems with categorizable reference-data space the power dissipation can be reduced to the value for one bank in the best case.

Proposed Architecture for Euclidean Distance Search

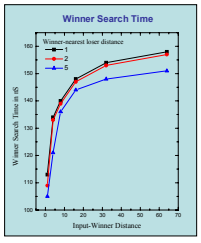


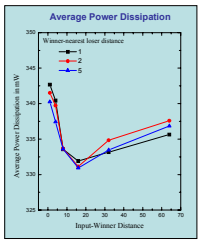
Features:

- **All unit comparators and word comparators calculate the distance between input pattern and stored reference patterns in parallel.**
- **Fast analog word comparison (e.g. comparison result encoded as static current-sink capability.)**
- **Analog squarer circuit for realizing the Euclidean distance function makes the system compact and easy to realize in hardware.**
- **Winner search circuitry scales only linear with the number of reference words R.**



Simulation of Minimum Euclidean Distance Search



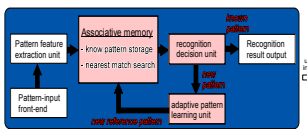


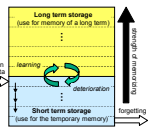
Distance Measure	Euclidean Distance 16, 5-bit Binaries
Winner-N, Loser Distance	1(One), 2 (Two) and 5(Five)
Reference Patterns	128
Technology	0.35 μm CMOS

- **Winner search time is between 105 and 160 nsec.**
- **Average power dissipation of the total circuit is less than 343 mW.**

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.

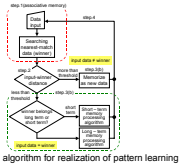
Automatic Associative Learning of Reference Pattern

Research Contents

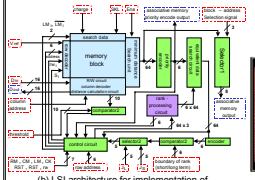
- Development of an LSI-architecture capable of pattern matching and pattern learning without teacher

Results

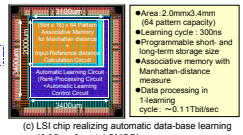
- Algorithm based on associative memory
- Realization of short- and long-term memory function
- LSI architecture and test-chip in 0.35um CMOS



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



(b) LSI architecture for implementation of the learning algorithm



- Area: 2.0mm x 3.4mm (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 : ~0.1 Tbit/sec

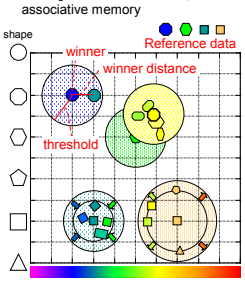
(c) LSI chip realizing automatic data-base learning (0.35um 3-metal CMOS)

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

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

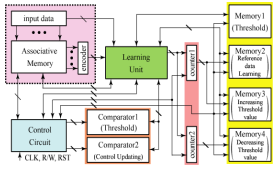
Reference data most similar to input data = winner

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 and VLSI Implementation of Reference-Pattern Learning and Optimization Algorithm



VLSI Architecture realizing reference-pattern optimization.

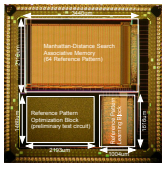


Photo of the test chip for associative memory with high-speed matching, learning and optimization of reference patterns. (0.35um 3-metal CMOS)

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

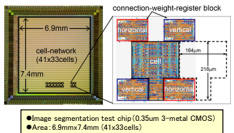
Low Power Cell-Network-Based Image Segmentation

Research Contents

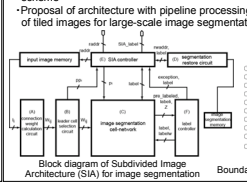
- Development of low power real-time color image segmentation chip architecture

Results

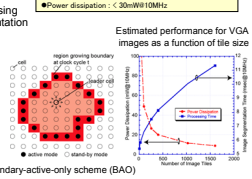
- Digital color image segmentation algorithm based on region growing approach
- Fully pixel parallel architecture based on cell network
- Low power operation with boundary-active-only scheme
- Proposal of architecture with pipeline processing of tiled images for large-scale image segmentation



- Image segmentation test chip (0.35um 3-metal CMOS)
- Area: 6.9mm x 7.4mm (41x33cells)
- Segmentation time : 23usec@10MHz (41x33pixels)
- Power dissipation : < 30mW@10MHz

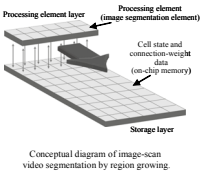


Block diagram of Subdivided Image Architecture (SIA) for image segmentation

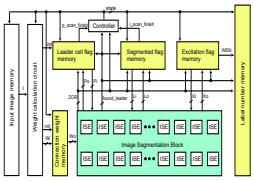


Boundary-active-only scheme (BAO)

Image-Scan Concept for Region-Growing Segmentation



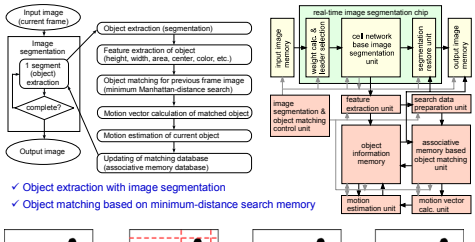
Conceptual diagram of image-scan video segmentation by region growing.



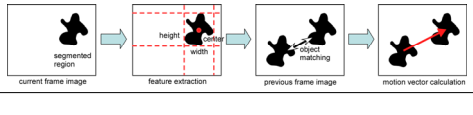
Embedded-memory-based VLSI architecture of image-scan video segmentation.

- Enables single-chip real-time segmentation of large-scale images
- Implementation with FPGA and SOC-ASIC possible
- Trade-off between hardware amount and segmentation speed possible

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
  - Capability for Hamming, Manhattan and Euclidean distance measures
- **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 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 with FPGA-based demonstration system