A Hierarchical Multiple-Chip K-means Processor System
for Real-Time Visual Learning and Perception
（視覚情報の実時間学習・認識のための階層型マルチチップ
K-means プロセッサシステム）

馬　奕涛

Benefiting from the continuous progress in semiconductor technology, billions of electronic devices can be integrated on a single VLSI processor chip, producing enormous computational powers in digital systems. As a result, the real-time learning and recognition function using visual information, one of the most critical sources of intelligence, are highly expected for time-critical applications such as automotive car control, video surveillance, robotic systems, and so forth. However, in spite of the success in logical and mathematical computation, the visual learning and visual recognition are both still very challenging task for traditional computer systems because of the high computational cost and a huge amount of processing data arising from pixel-wise operations carried out on images. A number of research approaches including those with software programs, with graphics processing units (GPUs) and with dedicated hardware system development have been widely investigated. However, they are usually each tuned to specific applications, and it is hard to find a real-time operating system that can handle both learning and recognition tasks on a general platform. At the same time, due to the limitation of hardware resources, there is a difficult trade-off between the high-speed performance and the large-scale data processing capability.

In contrast, the human brain can instantaneously achieve precise and flexible performance for visual information processing. In order to mimic the robust nature of the human brain, many bio-inspired approaches have been investigated for image learning and recognition. Based on the discovery of physiological research, a *brain-inspired* image representation algorithm using *directional edge information* was developed and its superior performance in image recognition has been demonstrated. Therefore, in this research, we have proposed an *image recognition brain model* and *an image learning brain model*, which both deal with a large quantity of directional-edge-based image feature vectors. And the final goal is to develop a dedicated hardware platform combining the two brain-inspired models, which can accomplish both real-time processing performance and large-scale data processing capability.

In this work, the image learning brain model has been realized by directly applying the K-means clustering algorithm to a large amount of 64-dimenssion edge-based image feature vectors which are extracted from every pixels of input large-size image by using a 64x64 scan window. And by clustering the generated feature vectors in the vector space, the noteworthy regions in original source

image are able to be automatically well segmented from the complicated background. The efficacy of this learning model for scene understanding and texture representation was demonstrated by software simulation running on a dual-core 2GHz CPU. On the other hand, the image recognition brain model simply utilizes the template matching algorithm to multidimensional feature vectors, which are generated from the input image by the same extraction method of the learning model, and a large number of image feature vectors pre-stored as past experience. The feature vector generation algorithm used in these two models had already been proposed and implemented in dedicated VLSI processors as the previous work in our laboratory. And the Manhattan distance computation is chosen as the similarity evaluation method in both the K-means algorithm and template matching algorithm.

For achieving the real-time performance of these two brain-inspired models, a dedicated hierarchical multiple-chip VLSI architecture having learning and recognition operating modes has been developed. As the most complex part of the proposed VLSI architecture, in order to achieve the real-time response of the K-means-based learning model, we employed an on-chip memory architecture and take the advantage of fully parallel distance calculation for all of the data. And for enhancing the processing capability, a binary-tree hierarchical multiple-chip architecture has been proposed, which allows the system to be extended unlimitedly to any scale by simply increasing the total number of dedicated chips with the same configuration. The key technology in the architecture is a multifrequency-driven pipeline scheme in which the delay time arising from the inter-chip data transmission has been effectively compensated for by intra-chip multiple distance computation operations, thus eliminating the inter-chip communication time loss, the bottleneck in achieving a large-throughout real-time performance. The proof-of-concept chip for a rank-4 multiple-chip system was designed in a 0.18 μm five-metal CMOS technology. The single chip operation at 100 MHz under a power supply voltage of 1.8V has been demonstrated by NanoSim simulation, and the pipeline calculation flow was demonstrated by measurement of the partial circuit (100MHz, 1.8V supply). The system can complete one K-means iteration cycle for partitioning 64-element learning vectors into 16 clusters in only 41.6 μs at 100 MHz. It is about 20,000 times faster than the same one K-means iteration by software processing running on a 2GHz dual-core general-purpose processor for XGA-size image segmentation. This performance is also significantly improved comparing with the other latest related works on GPUs and dedicated ASICs.

The only critical point of K-means-based learning is that the computation speed and clustering quality of K-means algorithm is very sensitive to initial cluster centers which are always randomly chosen from the learning sample vectors at the beginning. In order to automatically determine the number and values of the initial cluster centers and further enhance the performance of the K-means-base learning model, a hardware-friendly adaptive K-means algorithm has been proposed. In this adaptive algorithm, K-means clustering for the same vector data set is repeated while the

number of initial cluster centers (K) is increasing from one to a given maximum value. And in order to find the optimum value of K, the performance for every different chosen cluster number is evaluated using the *Variance Ratio Criterion (VRC)* function which is a popular information criterion for model selection in statistics. And for every set number of K, the exact values of initial cluster centers are calculated using a *farthest selection* method. In the method, the Manhattan distances among all of the learning sample vectors are calculated. And the values of initial cluster centers are chosen to be as far as possible from each other, thus making the distance among cluster centers largest. The efficacy of this adaptive K-means learning algorithm for image segmentation was demonstrated by MATLAB simulation running on a dual-core 3GHz CPU. Because of the Manhattan distance computations for all data, the computational cost of the *farthest selection* algorithm depends on the number of learning sample vectors, which makes this proposed adaptive K-means algorithm very computational expensive for software approach. On the other hand, however, it is very suitable to be combined into the existing dedicated architecture for K-means-based learning brain model, which can accelerate the algorithm by employing the massively parallel for the Manhattan distance calculation.

Therefore, a new architecture has been proposed as a variant of the binary-tree hierarchical multiple-chip K-means architecture, which has only two additional circuitries. Firstly, a VRC estimation circuitry has been added for determining the optimum number of K-means initial cluster centers. Then, a hierarchical winner-take-all (WTA) circuitry has been added to carry out the largest one from the Manhattan distance results, which is necessary to determine the exact values of given number of K-means initial cluster centers. It should be noted that, in this new architecture, the WTA circuit is also used for the minimum searching function of the template matching recognition model, which makes it possible to achieve real-time and large-scale performance of visual recognition on the same hardware platform. The operations of additional circuitries were verified by NanoSim hardware simulation.