

A Real-Time VLSI Recognition System With an On-Chip Adaptive K-Means Learning Algorithm

Zuoxun Hou^{1,4}, Yitao Ma², Hongbo Zhu³ and Tadashi Shibata⁴

¹ Inst. of Artificial Intelligence and Robotics, Xi'an Jiaotong Univ. Xi'an, China

² Center for Interdisciplinary Res. (CIR), Tohoku Univ. Sendai, Japan

³ VLSI Design & Education Center (VDEC), Univ. of Tokyo, Tokyo, Japan

⁴ Dept. of Electrical Engineering and Information System, Univ. of Tokyo, Tokyo, Japan
E-mail: houzuoxun1@gmail.com, shibata@ee.t.u-tokyo.ac.jp

1. Introduction

Real time recognition of objects or events is now becoming increasingly important in a number of intelligent systems such as unmanned surveillance systems, autonomous vehicle control, robotics, and so forth. Recognition results must be available in real time with minimum latency at very low power dissipation, and software solutions running on general purpose computers often fail to fulfill such requirements. In this paper, we will present a VLSI hardware solution not only for real time recognition, but also for real time learning of a large number of samples.

Recognition is understood as a classification problem as illustrated in Fig. 1. In supervised learning, a number of samples are stored in the system as templates for each person with a proper class label such as Chaplin, Einstein etc., and when an unknown facial image is presented, the system searches for the best match image and outputs its class label as the recognition result. Usually such pattern matching is carried out using feature vectors representing original images in a compact format based on some algorithms like the one described in [1]. The VLSI chip developed in the present work can carry out such a search in fully parallel processing and can yield the recognition result in as short as $3\mu\text{s}$ for 64-dimension feature vectors at 25MHz of operation.

In the case of unsupervised learning, a number of sample images are presented to the system without any class labels, and the system must categorize them by itself. Namely, the system must autonomously divide the entire samples into four clusters as in the example of Fig. 1. For such processing, K-means clustering algorithm is very widely used [2]. However, since it requests a number of distance calculations and iterations, real time response is very difficult to achieve in software solutions. Therefore, VLSI architectures dedicated to the K-means algorithm were developed, and VLSI chip implementation [3] and FPGA implementation [4] were reported. One of the algorithmic problems of K-means is that K , the number of clusters, must be determined in advance. Therefore, if we set $K=5$ in the example of Fig. 1, the samples are over segmented. In the present work, not only the K-means clustering algorithm but also an autonomous K-number determining mechanism have been implemented on the same chip, in addition to the recognition function in the supervised learning mode. As a result, a very versatile VLSI architecture for

recognition as well as for learning has been established.

The architecture was implemented on Altera cyclone II FPGA (EP2C70896C6) chip and the real time K-means classification performance was experimentally demonstrated using COIL-100 (Columbia University Object Image Library) database.

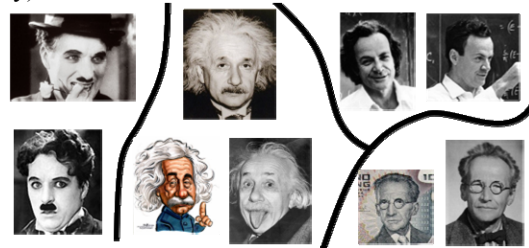


Fig. 1. Recognition as a classification problem.

2. Adaptive K-means Learning Algorithm

The quality of classification in the unsupervised learning is characterized by both small intra-class distances and large inter-class distances among samples. This means the samples in the same class (intra-class) are very similar to each other, while the samples in different classes (inter-class) are significantly different. Therefore the quality of K-means classification results can be evaluated using the Variance Ratio Criterion (VRC) [5] defined as follows:

$$VRC = \frac{SSb}{SSw} \times \frac{(N-K)}{(K-1)} \quad (1)$$

$$SSb = \sum_{i=1}^K SN_i \times |Centroid_i - GG| \quad (2)$$

$$SSw = \sum_{i=1}^K \sum_{j=1}^{SN_i} |Sample_{ij} - Centroid_i| \quad (3)$$

Here N is the total number of samples; K is the number of classes. i and j represent class and sample indices, respectively. GG is the global gravity center, $Centroid_i$ is the gravity center of class i . SN_i is the number of samples in class i . SSb reflects the average inter-class distance, while SSw the average intra-class distance. The $(N-K)/(K-1)$ term is introduced to normalize the value with respect to K . The adaptive K-means learning algorithm determines the proper K value by computing the VRC for $K=2, 3 \dots$ sequentially until it reaches the absolute or local maximum.

3. Hardware Architecture

Fig. 2(a) shows the system architecture. Dedicated par-

allel processing circuitries are developed for classification. All these circuitries are controlled by the system controller in which a finite state machine (FSM) is used as shown in Fig. 2(c). At the beginning, the FSM chooses the GG and the sample vector farthest from the GG as two initial centroids. Then start to cluster with $K=2$ and evaluate the VRC to decide whether to start a new iteration for $K=3$ or not. The VRC is calculated after each iteration to see whether a proper K value has been reached or not. As shown in Fig. 2(a), the feature vectors of all samples are preserved in the SRAM. There are two separate paths for feature vectors to be processed to generate SSb and SSw . In the first path, feature vectors are directly transferred to CIDM, and processed by LA and GCU to generate each centroid. At the same time, VRC (Fig. 2(b)) unit calculates the SSb in eq. (2). In the second path, DUs calculate the distance between each feature vector and each centroid, and update the contents in the cluster register (CID Reg) and the distance register (Distance Reg) when the distance is smaller than the stored value. After this, the system accumulates the values of distance registers to generate the SSw by LA and GCU. With SSb and SSw , the system calculates the VRC according to the present K . Here the term $(N-K)$ is approximated as N because $N \gg K$. If the present VRC value is smaller than the last VRC value, which means last VRC value reaches the local maximum and the process is terminated; else it selects the farthest vector as a new centroid using WTA unit and continues clustering for $K+1$.

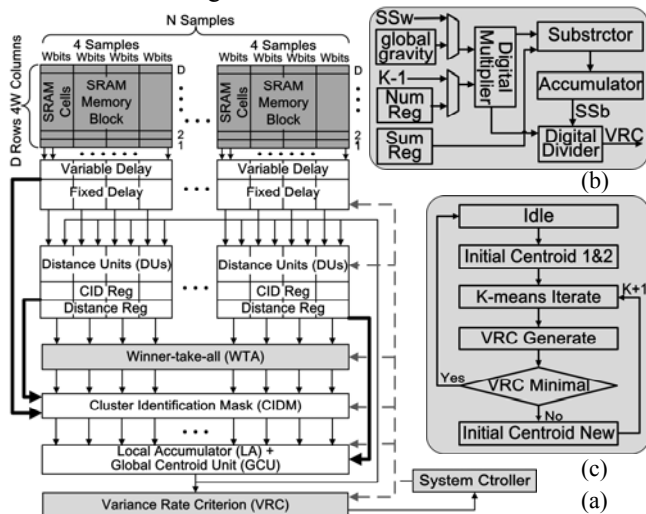


Fig. 2. Hardware architecture of the system: (a) global structure; (b) hardware structure of VRC; (c) flow chart of system controller.

4. Experimental Results

64 sample images (in 8 classes and 8 samples per class) selected from the COIL-100 database are used for verifying the performance of this system. The resolution of sample images is resampled to 64×64 pixels for generating the

PPED vectors [1]. Fig. 3 shows representative images from each class.

Fig. 4 shows the measured waveforms from oscilloscope. When clustering starts, the system executes the FSM described in Fig. 2(c) for $K=2, 3, \dots$, up to $K=9$. Then the system stops; which indicates that $K=8$ has reached the maximum VRC value. Fig. 5 shows measurement results captured using the embedded Altera SignalTap Logic Analyzer. The VRC value is calculated for each K ; and CID_Reg0 and CID_Reg63 present the cluster label change of sample0 and sample63 with the target class number K respectively. It is worth noting that VRC reaches its first local maximum when $K=2$, which is much smaller than the second one when $K=8$. To trade off between the speed performance and the global optimum value, two local maxima in limited K values are searched, and the larger one is chosen as the final result. The system clock is 25 MHz, and the total processing time is less than 0.3ms. Thanks to the scalable parallel processing structure, this architecture can be simply extended for processing larger number of samples with same computation time.

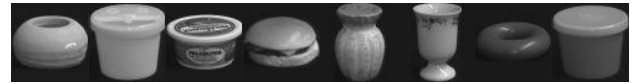


Fig. 3. Test samples belonging to 8 classes.

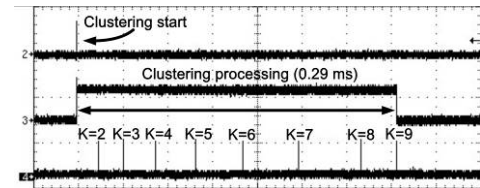


Fig. 4. Measured waveforms from oscilloscope.

5. Conclusion

A VLSI recognition system equipped with an adaptive K-means learning capability has been developed and the operation has been experimentally demonstrated. The eight category images taken from COIL-100 data base have been successfully classified into eight categories in less than 0.3 ms using the autonomous K-value determination function installed in the system.

References

- [1] M. Yagi and T. Shibata, *IEEE Trans. Neural Networks* **14**, 1144 (2003).
- [2] J. MacQueen, *Proc. 5th Berkley Symposium on Mathematical Statistics and Probability*, **1**, 164 (1967).
- [3] Y. Ma and T. Shibata, *Jpn. J. Appl. Phys.* **49**, 04DE08 (2010).
- [4] T. Maruyama, *ICPR*, **2**, 816 (2006).
- [5] L. Vendramin, R. Campello and E. Hruschka, *SIAM2009*, 733

Name	-1024	-512	0	512	1024	1536	2048	2560	3072	3584	4096	4608	5120	5632	6144	6656	7168
Clustering start	[Signal transition]																
Clustering processing	[Signal transition]																
⊕ K value	9h	2h	3h	4h	5h	6h	7h	8h	9h								
⊕ VRC	017h	000h	009h	008h	009h	00Ah	00Dh	019h	017h								
⊕ CID_Reg0	6h																
⊕ CID_Reg63	7h	10h															

Fig. 5. Measured waveforms from Altera SignalTap Logic Analyzer.