Chapter from the book *Principal Component Analysis*
Downloaded from: http://www.intechopen.com/books/principal-component-analysis

Interested in publishing with InTechOpen?
Contact us at book.department@intechopen.com
FPGA Implementation for GHA-Based Texture Classification

Shiow-Jyu Lin\textsuperscript{1,2}, Kun-Hung Lin\textsuperscript{1} and Wen-Jyi Hwang\textsuperscript{1}

\textsuperscript{1}Department of Computer Science and Information Engineering, National Taiwan Normal University and
\textsuperscript{2}Department of Electronic Engineering, National Ilan University, Taiwan

1. Introduction

Principal components analysis (PCA) (Alpaydin, 2010; Jolliffe, 2002) is an effective unsupervised feature extraction algorithm for pattern recognition, classification, computer vision or data compression (Bravo et al., 2010; Zhang et al., 2006; Kim et al., 2005; Liying & Weiwei, 2009; Pavan et al., 2007; Qian & James, 2008). The goal of PCA is to obtain a compact and accurate representation of the data that reduces or eliminates statistically redundant components. Basic approaches for PCA involve the computation of the covariance matrix and the extraction of eigenvalues and eigenvectors. A drawback of the basic approaches is the high computational complexity and large memory requirement for data with high vector dimension. Therefore, these approaches may not be well suited for real time applications requiring fast feature extraction.

A number of fast algorithms (Dogaru et al., 2004; El-Bakry, 2006; Gunter et al., 2007; Sajid et al., 2008; Sharma & Paliwal, 2007) have been proposed to reduce the computation time of PCA. However, only moderate acceleration can be achieved because most of these algorithms are based on software. Although hardware implementation of PCA and its variants are possible, large storage size and complicated circuit control management are usually necessary. The PCA hardware implementation may therefore be possible only for small dimensions (Boonkumklao et al., 2001; Chen & Han, 2009).

An alternative for the PCA implementation is to use the generalized Hebbian algorithm (GHA) (Haykin, 2009; Oja, 1982; Sanger, 1989). The principal computation by the GHA is based on an effective incremental updating scheme for reducing memory utilization. Nevertheless, slow convergence of the GHA (Karhunen & Joutsensalo, 1995) is usually observed. A large number of iterations therefore is required, resulting in long computational time for many GHA-based algorithms. The hardware implementation of GHA has been found to be effective for reducing the computation time. However, since the number of multipliers in the circuit grows with the dimension, the circuits may be suitable only for PCA with small dimensions. Although analog GHA hardware architectures (Carvajal et al., 2007; 2009) can be used to lift the constraints on the vector dimensions, these architectures are difficult to be directly used for digital devices.
In light of the facts stated above, a digital GHA hardware architecture capable of performing fast PCA for large vector dimension is presented. Although large amount of arithmetic computations are required for GHA, the proposed architecture is able to achieve fast training with low area cost. The proposed architectures can be divided into three parts: the synaptic weight updating (SWU) unit, the principal components computing (PCC) unit, and memory unit. The memory unit is the on-chip memory storing synaptic weight vectors. Based on the synaptic weight vectors stored in the memory unit, the SWU and PCC units are then used to compute the principal components and update the synaptic weight vectors, respectively.

In the SWU unit, one synaptic weight vector is computed at a time. The results of precedent weight vectors will be used for the computation of subsequent weight vectors for expediting training speed. In addition, the computation of different weight vectors shares the same circuit for lowering the area cost. Moreover, in the PCC unit, the input vectors are allowed to be separated into smaller segments for the delivery over data bus with limited width. Both the SWU and PCC units can also operate concurrently to further enhance the throughput.

To demonstrate the effectiveness of the proposed architecture, a texture classification system on a system-on-programmable-chip (SOPC) platform is constructed. The system consists of the proposed architecture, a softcore NIOS II processor (Altera Corp., 2010), a DMA controller, and a SDRAM. The proposed architecture is adopted for finding the PCA transform by the GHA training, where the training vectors are stored in the SDRAM. The DMA controller is used for the DMA delivery of the training vectors. The softcore processor is only used for coordinating the SOPC system. It does not participate the GHA training process. As compared with its software counterpart running on Intel i7 CPU, our system has significantly lower computational time for large training set. All these facts demonstrate the effectiveness of the proposed architecture.

Fig. 1. The neural model for the GHA.

2. Preliminaries

Figure 1 shows the neural model for GHA, where \( x(n) = [x_1(n), ..., x_m(n)]^T \); and \( y(n) = [y_1(n), ..., y_p(n)]^T \) are the input and output vectors to the GHA model, respectively. The output
vector $y(n)$ is related to the input vector $x(n)$ by

$$y_j(n) = \sum_{i=1}^{m} w_{ji}(n)x_i(n), \quad (1)$$

where the $w_{ji}(n)$ stands for the weight from the $i$-th synapse to the $j$-th neuron at iteration $n$. Each synaptic weight vector $w_j(n)$ is adapted by the Hebbian learning rule:

$$w_{ji}(n+1) = w_{ji}(n) + \eta [y_j(n)x_i(n) - y_j(n) \sum_{k=1}^{j} w_{kj}(n)y_k(n)], \quad (2)$$

where $\eta$ denotes the learning rate. After a large number of iterative computation and adaptation, $w_j(n)$ will asymptotically approach to the eigenvector associated with the $j$-th principal component $\lambda_j$ of the input vector, where $\lambda_1 > \lambda_2 > ... > \lambda_p$. To reduce the complexity of computing implementation, eq.(2) can be rewritten as

$$w_{ji}(n+1) = w_{ji}(n) + \eta y_j(n)[x_i(n) - \sum_{k=1}^{i} w_{ki}(n)y_k(n)]. \quad (3)$$

A more detailed discussion of GHA can be found in (Haykin, 2009; Sanger, 1989)

3. The proposed GHA architecture

![Fig. 2. The proposed GHA architecture.](image)

As shown in Figure 2, the proposed GHA architecture consists of three functional units: the memory unit, the synaptic weight updating (SWU) unit, and the principal components computing (PCC) unit. The memory unit is used for storing the current synaptic weight vectors. Assume the current synaptic weight vectors $w_j(n), j = 1, ..., p$, are now stored in the memory unit. In addition, the input vector $x(n)$ is available. Based on $x(n)$ and $w_j(n), j = 1, ..., p$, the goal of PCC unit is to compute output vector $y(n)$. Using $x(n), y(n)$
and \( w_j(n), j = 1, ..., p \), the SWU unit produces the new synaptic weight vectors \( w_j(n + 1), j = 1, ..., p \). It can be observed from Figure 2 that the new synaptic weight vectors will be stored back to the memory unit for subsequent training.

### 3.1 SWU unit

The design of SWU unit is based on eq.(3). Although the direct implementation of eq.(3) is possible, it will consume large hardware resources. To further elaborate this fact, we first see from eq.(3) that the computation of \( w_{ji}(n + 1) \) and \( w_{ri}(n + 1) \) shares the same term \( \sum_{k=1}^{r} w_{ki}(n)y_k(n) \) when \( r \leq j \). Consequently, independent implementation of \( w_{ji}(n + 1) \) and \( w_{ri}(n + 1) \) by hardware using eq.(3) will result in large hardware resource overhead.

Fig. 3. The hardware implementation of eqs.(5) and (6).

To reduce the resource consumption, we first define a vector \( z_{ji}(n) \) as

\[
z_{ji}(n) = x_i(n) - \sum_{k=1}^{j} w_{ki}(n)y_k(n), j = 1, ..., p, \tag{4}
\]

and \( z_j(n) = [z_{j1}(n), ..., z_{jm}(n)]^T \). Integrating eq.(3) and (4), we obtain

\[
w_{ji}(n + 1) = w_{ji}(n) + \eta y_j(n)z_{ji}(n), \tag{5}
\]

where \( z_{ji}(n) \) can be obtained from \( z_{(j-1)i}(n) \) by

\[
z_{ji}(n) = z_{(j-1)i}(n) - w_{ji}(n)y_j(n), j = 2, ..., p, \tag{6}
\]

When \( j = 1 \), from eq.(4) and (6), it follows that

\[
z_{0i}(n) = x_i(n). \tag{7}
\]

Figure 3 depicts the hardware implementation of eqs.(5) and (6). As shown in the figure, the SWU unit produces one synaptic weight vector at a time. The computation of \( w_j(n + 1) \), the \( j \)-th weight vector at the iteration \( n + 1 \), requires the \( z_{j-1}(n) \), \( y(n) \) and \( w_j(n) \) as inputs.
Fig. 4. The architecture of each module in SWU unit.

In addition to \( w_j(n+1) \), the SWU unit also produces \( z_j(n) \), which will then be used for the computation of \( w_{j+1}(n+1) \). Hardware resource consumption can then be effectively reduced.

One way to implement the SWU unit is to produce \( w_j(n+1) \) and \( z_j(n) \) in one shot. In SWU unit, \( m \) identical modules may be required because the dimension of vectors is \( m \). Figure 4 shows the architecture of each module. The area cost of the SWU unit then will grow linearly with \( m \). To further reduce the area cost, each of the output vectors \( w_j(n+1) \) and \( z_j(n) \) are separated into \( b \) segments, where each segment contains \( q \) elements. The SWU unit only computes one segment of \( w_j(n+1) \) and \( z_j(n) \) at a time. Therefore, it will take \( b \) clock cycles to produce complete \( w_j(n+1) \) and \( z_j(n) \).

Let

\[
\hat{w}_{j,k}(n) = [w_j(k-1)q+1(n), ..., w_j(k-1)q+q(n)]^T, k = 1, ..., b. \tag{8}
\]

and

\[
\hat{z}_{j,k}(n) = [z_j(k-1)q+1(n), ..., z_j(k-1)q+q(n)]^T, k = 1, ..., b. \tag{9}
\]

be the \( k \)-th segment of \( w_j(n) \) and \( z_j(n) \), respectively. The computation \( w_j(n) \) and \( z_j(n) \) take \( b \) clock cycles. At the \( k \)-th clock cycle, \( k = 1, ..., b \), the SWU unit computes \( \hat{w}_{j,k}(n+1) \) and \( \hat{z}_{j,k}(n) \). Because each of \( \hat{w}_{j,k}(n+1) \) and \( \hat{z}_{j,k}(n) \) contains only \( q \) elements, the SWU unit consists of \( q \) identical modules. The architecture of each module is also shown in Figure 4. The SWU unit can be used for GHA with different vector dimension \( m \). As \( m \) increases, the area cost therefore remains the same at the expense of large number of clock cycles \( b \) for the computation of \( \hat{w}_{j,k}(n+1) \) and \( \hat{z}_{j,k}(n) \).

Figures 5, 6 and 7 show the operation of the \( q \) modules. For the sake of simplicity, the computation of the first weight vector \( w_1(n+1) \) (i.e., \( j = 1 \)) and the corresponding \( z_1(n) \) are considered in the figures. Based on eq.(7), the input vector \( z_{j-1}(n) \) is actually the training
vector $x(n)$, which is also separated into $b$ segments, where the $k$-th segment is given by

$$
\hat{z}_{0,k}(n) = [x_{(k-1)q+1}(n), ..., x_{(k-1)q+q}(n)]^T, k = 1, ..., b.
$$

(10)

They are then multiplexed to the $q$ modules. The $\hat{z}_{0,1}(n)$ and $\hat{w}_{1,1}(n)$ are used for the computation of $\hat{z}_{1,1}(n)$ and $\hat{w}_{1,1}(n+1)$ in Figure 5. Similarly, the $\hat{z}_{0,k}(n)$ and $\hat{w}_{1,k}(n)$ are
Fig. 7. The operation of SWU unit for computing the $b$-th segment of $w_1(n+1)$.

Fig. 8. The operation of SWU unit for computing the first segment of $w_2(n+1)$.

used for the computation of $\hat{z}_{1,k}(n)$ and $w_{1,k}(n+1)$ in Figures 6 and 7 for $k = 2$ and $k = b$, respectively.

After the computation of $w_1(n+1)$ is completed, the vector $z_1(n)$ is available as well. The vector $z_1(n)$ is then used for the computation of $w_2(n+1)$. Figure 8 shows the computation of the first segment of $w_2(n+1)$ (i.e., $\hat{w}_{2,1}(n+1)$) based on the first segment of $z_1(n)$ (i.e., $\hat{z}_{1,1}(n)$). The same process proceeds for the subsequent segments until the computation of
the entire vectors $\mathbf{w}_2(n+1)$ and $\mathbf{z}_2(n)$ are completed. The vector $\mathbf{z}_2(n)$ is then used for the computation of $\mathbf{w}_3(n+1)$. The weight vector updating process at the iteration $n+1$ will be completed until the SWU unit produces the weight vector $\mathbf{w}_p(n+1)$.

---

**Fig. 9.** The architecture of Buffer A in memory unit.

**Fig. 10.** The architecture of Buffer B in memory unit.
3.2 Memory unit

The memory unit contains three buffers: Buffer A, Buffer B and Buffer C. As shown in Figure 9, Buffer A stores training vector $x(n)$. It consists of $q$ sub-buffers, where each sub-buffer contains $b$ elements. All the sub-buffers are connected to the SWU and PCC units.

The architecture of Buffer B is depicted in Figure 10, which holds the values of $z_j(n)$. Each segment of $z_j(n)$ computed from SWU unit is stored in Buffer B. After all the segments are produced, the Buffer B then deliver the segments of $z_j(n)$ to SWU unit in the first-in-first-out (FIFO) fashion.

The Buffer C is used for storing the synaptic weight vectors $w_j(n), j = 1, \ldots, p$. It is a two-port RAM for reading and writing weight vectors, as revealed in Figure 11. The address for the RAM is expressed in terms of indices $j$ and $i$ for reading or writing the $i$-th segment of the weight vector $w_j(n)$.

3.3 PCC unit

The PCC operations are based on eq.(1). Therefore, the PCC unit of the proposed architecture contains adders and multipliers. Figure 12 shows the architecture of PCC. The training vector $x(n)$
x(n) and synaptic weight vector w_j(n) are obtained from the Buffer A and Buffer C of the memory unit, respectively. When both x(n) and w_j(n) are available, the proposed PCC unit then computes y_j(n). Note that, after the y_j(n) is obtained, the SWU unit can then compute w_j(n + 1). Figure 13 reveals the timing diagram of the proposed architecture. It can be observed from Figure 13 that both the y_{j+1}(n) and w_j(n + 1) are computed concurrently. The throughput of the proposed architecture can then be effectively enhanced.

3.4 SOPC-based GHA training system

The proposed architecture is used as a custom user logic in a SOPC system consisting of softcore NIOS CPU (Altera Corp., 2010), DMA controller and SDRAM, as depicted in Figure 14. All training vectors are stored in the SDRAM and then transported to the proposed circuit via the Avalon bus. The softcore NIOS CPU runs on a simple software to coordinate different components, including the proposed custom circuit in the SOPC. The proposed circuit operates as a hardware accelerator for GHA training. The resulting SOPC system is able to perform efficient on-chip training for GHA-based applications.

4. Experimental results

This section presents some experimental results of the proposed architecture applied to texture classification. The target FPGA device for all the experiments in this paper is Altera Cyclone III (Altera Corp., 2010). The design platform is Altera Quartus II with SOPC Builder and NIOS II IDE. Two sets of textures are considered in the experiments. The first set of textures, shown in Figure 15, consists of three different textures. The second set of textures is revealed in Figure 16, which contains four different textures. The size of each texture in Figures 15 and 16 is 320 × 320.

In the experiment, the principal component based k nearest neighbor (PC-kNN) rule is adopted for texture classification. Two steps are involved in the PC-kNN rule. In the first step, the GHA is applied to the input vectors to transform m dimensional data into p principal components. The synaptic weight vectors after the convergence of GHA training are adopted to span the linear transformation matrix. In the second step, the kNN method is applied to the principal subspace for texture classification.
Fig. 14. The SOPC system for implementing GHA.

Fig. 15. The first set of textures for the experiments.

Figures 17 and 18 show the distribution of classification success rates (CSR) of the proposed architecture for the texture sets in Figures 15 and 16, respectively. The classification success rate is defined as the number of test vectors which are successfully classified divided by the total number of test vectors. The number of principal components is $p = 4$. The vector dimension is $m = 16 \times 16$. The distribution is based on 20 independent GHA training processes. The distribution of the architecture presented in (Lin et al., 2011) with the same $p$ is also included for comparison purpose. The vector dimension for (Lin et al., 2011) is $m = 4 \times 4$. 
Fig. 16. The second set of textures for the experiments.

Fig. 17. The distribution of CSR of the proposed architecture for the texture set in Figure 15.

It can be observed from Figures 17 and 18 that the proposed architecture has better CSR. This is because the vector dimension of the proposed architecture is higher than that in (Lin et al., 2011). Spatial information of textures therefore is more effectively exploited for improving CSR by the proposed architecture. In fact, the vector dimension in the proposed architecture is $m = 16 \times 16$. The proposed architecture is able to implement hardware GHA for larger
vector dimension because the area cost of the SWU unit in the architecture is independent of vector dimension. By contrast, the area cost of the SWU unit in (Lin et al., 2011) grows with the vector dimension. Therefore, only smaller vector dimension (i.e., \( m = 4 \times 4 \)) can be implemented.

To further elaborate these facts, Tables 1 and 2 show the hardware resource consumption of the proposed architecture and the architecture in (Lin et al., 2011) for vector dimensions \( m = 4 \times 4 \) and \( m = 16 \times 16 \). Three different area costs are considered in the table: logic elements (LEs), embedded memory bits, and embedded multipliers. It can be observed from Tables 1 and 2 that given the same \( m = 4 \times 4 \) and the same \( p \), the proposed architecture consumes significantly less hardware resources as compared with the architecture in (Lin et al., 2011). Although the area costs of the proposed architecture increase as \( m \) becomes \( 16 \times 16 \), as shown in Table 1, they are only slightly higher than those of (Lin et al., 2011) in Table 2. The proposed architecture therefore is well suited for GHA training with larger vector dimension due to better spatial information exploitation.

<table>
<thead>
<tr>
<th>( p )</th>
<th>Proposed GHA with ( m = 4 \times 4 )</th>
<th>Proposed GHA with ( m = 16 \times 16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LEs</td>
<td>Memory Bits</td>
</tr>
<tr>
<td>3</td>
<td>3942</td>
<td>1152</td>
</tr>
<tr>
<td>4</td>
<td>4097</td>
<td>1152</td>
</tr>
<tr>
<td>5</td>
<td>4394</td>
<td>1280</td>
</tr>
<tr>
<td>6</td>
<td>4686</td>
<td>1280</td>
</tr>
<tr>
<td>7</td>
<td>4988</td>
<td>1280</td>
</tr>
</tbody>
</table>

Table 1. Hardware resource consumption of the proposed GHA architecture for vector dimensions \( m = 4 \times 4 \) and \( m = 16 \times 16 \).
Table 2. Hardware resource consumption of the GHA architecture (Lin et al., 2011) for vector dimension $m = 4 \times 4$.

<table>
<thead>
<tr>
<th>$p$</th>
<th>LEs</th>
<th>Memory Bits</th>
<th>Embedded Multipliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>22850</td>
<td>0</td>
<td>204</td>
</tr>
<tr>
<td>4</td>
<td>31028</td>
<td>0</td>
<td>272</td>
</tr>
<tr>
<td>5</td>
<td>38261</td>
<td>0</td>
<td>340</td>
</tr>
<tr>
<td>6</td>
<td>45991</td>
<td>0</td>
<td>408</td>
</tr>
<tr>
<td>7</td>
<td>53724</td>
<td>0</td>
<td>476</td>
</tr>
</tbody>
</table>

Fig. 19. The CPU time of the NIOS-based SOPC system using the proposed architecture as the hardware accelerator for various numbers of training iterations with $p = 4$.

Figure 19 shows the CPU time of the NIOS-based SOPC system using the proposed architecture as the hardware accelerator for various numbers of training iterations with $p = 4$. The clock rate of NIOS CPU in the system is 50 MHz. The CPU time of the software counterpart also is depicted in the Figure 19 for comparison purpose. The software training is based on the general purpose 2.67-GHz Intel i7 CPU. It can be clearly observed from Figure 16 that the proposed hardware architecture attains high speedup over its software counterpart. In particular, when the number of training iterations reaches 1800, the CPU time of the proposed SOPC system is 1861.3 ms. By contrast, the CPU time of Intel i7 is 13860.3 ms. The speedup of the proposed architecture over its software counterpart therefore is 7.45.

5. Concluding remarks

Experimental results reveal that the proposed GHA architecture has superior speed performance over its software counterparts. In addition, the architecture is able to attain higher classification success rate for texture classification as compared with other existing
GHA architectures. The architecture also has low area cost for PCA analysis with high vector dimension. The proposed architecture therefore is an effective alternative for on-chip learning applications requiring low area cost, high classification success rate and high speed computation.

6. References


http://www.altera.com/literature/lit-nio2.jsp


This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as image processing, biometric, face recognition and speech processing. It also includes the core concepts and the state-of-the-art methods in data analysis and feature extraction.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: