REAL TIME ON-CHIP SEQUENTIAL ADAPTIVE PRINCIPAL COMPONENT ANALYSIS FOR DATA FEATURE EXTRACTION AND IMAGE COMPRESSION

Tuan A Duong

Bio-Inspired Technologies and Systems Group

Avionic Systems and Technology Division

Jet Propulsion Laboratory

California Institute of Technology

Pasadena, CA 91109

Abstract:

In this paper, we present a new, simple, and optimized hardware architecture sequential learning technique for adaptive Principal Component Analysis (PCA) which will "help optimize the hardware implementation" in VLSI and to overcome the difficulties of the traditional gradient descent in learning convergence and hardware implementation.

To demonstrate the performance of our algorithm, we study two cases: data feature extraction and image compression. In this study, we show that our learning approach can extract the principal components from the given data set as compatible as PCA using deterministic approach from MATLAB. However, the innovation of our approach is simple and easy to implement in hardware that can optimize hardware requirements for System-On-A-Chip approach. In addition, it demonstrates that our approach is more advanced in hardware implementation when compared with state of the art for gradient descent technique.

I. INTRODUCTION

Principal Component Analysis (PCA) is a second order statistical approach, which has been used to extract the features of data set [1] or perform data reduction (compression) [2,3]. Specially, when data set is, redundant and overwhelming large, PCA is very effective linear technique as a preprocessing step to extract data features and to cluster data for classification. It can play as optimal linear transform known as Kahunen-Louvre (LK) for data compression.

To obtain the principal component vectors, traditionally the covariance matrix is calculated then eigen values are obtained, and corresponding to each eigen value, a component (eigen) vector is found. This procedure is complicated and computationally intensive thereby making it restrictive to apply for real world applications such as data compression and data extraction. Moreover, the PCA hardware implementation for real time application becomes even more challenging.

To get over the hurdles from the traditional PCA technique, the simple sequential PCA techniques are introduced [4-10]. These techniques are based on learning approach to obtain sequentially principal component vectors. Some works in PCA are reported using Hebbian or anti-Hebbian learning [4-5] and gradient-based learning [6-11]. There are several reports that are successful in using PCA for data reduction and detection [12-13]. Most of the works are software-based due to the complication of the hardware requirements.

For the gradient descent technique, it is straight forward to compare with others e.g. steepest decent, conjugate gradient, or Newton's second order, but it still poses some difficulties: learning convergence with minor component vectors and further, it is still complicated to implement in hardware.

We introduce an innovative *dominant element* based gradient descent technique to simplify the system architecture and to reduce computational intensity as required for gradient descent. This simplification will ensure, at least, the same quality convergence and require much less hardware if implemented as opposed gradient descent technique.

Due to the collapse of the energy function when each component vector is extracted and removed consecutively from the energy function; hence the next component vector faces challenge to be obtained due to the very small attractor to compare with the previous ones. Moreover, we introduce the innovative *dynamic initial learning rate* to compensate for the energy lost when the previous components are removed from the energy function.

In this paper, we provide the new learning algorithm named *dominant element based gradient descent and dynamic initial learning rate* (DOGEDYN). This DOGEDYN is used to study feature extraction and data compression to demonstrate its superiority as opposed to gradient descent (GED), dominant element based gradient descent (DOGED), gradient descent with dynamic initial learning rate (GEDYN). Finally, a simple hardware architecture is proposed to solve hyperspectral application for full parallelism in computation to obtain real time computation capability.

II. MATHEMATICAL FOUNDATION

We adapt the objective function [11] below:

$$J(w)=\sum_{i=1}^{m}J_i(w_i)=\sum_{i=1}^{m}\sum_{t=1}^{k}\left|x_t-w_iw_i^Tx_t\right|^2 \quad (1)$$

where m is the number of principal components and k is the number of measurement vectors. x_t is a measured vector at time t and w_i is the i^{th} principal vector (or eigen vector).

With

$$J_i(w_i)=\sum_{t=1}^{k}\left|y_t-w_iw_i^Ty_t\right|^2$$

And $y_t=x_t-\sum_{j=1}^{i-1}w_jw_j^Tx_t$ (2)

PCA learning approach

From equation (2), the learning algorithm can be processed sequentially for each principal vector that is based on the gradient descent as follows:

$$\Delta w_{ij}=-\frac{\partial\Psi_i}{\partial w_{ij}}=-\frac{\partial(\left|y_i-w_iw_i^Ty_i\right|^2)}{\partial w_{ij}} \quad (3)$$

From equation (3), only the dominant element [14] (full analysis will be provided in the long paper) is used; the weight update can be obtained as follows:

$$w_{ij}^{new}=w_{ij}^{old}+\zeta\Delta w_{ij}=w_{ij}^{old}+\zeta\varepsilon_{ij}(w_i^Ty_i+w_{ij}y_{ij}) \quad (4)$$

Where $\zeta=\frac{E_0}{E_{i-1}}$

E_0 is the initial energy when the network starts learning and E_{i-1} the energy of the $(i-1)^{th}$ extracted principal has.

III. HARDWARE ARCHITECTURE

From equation (4), the learning architecture is realized as in Figure 3.

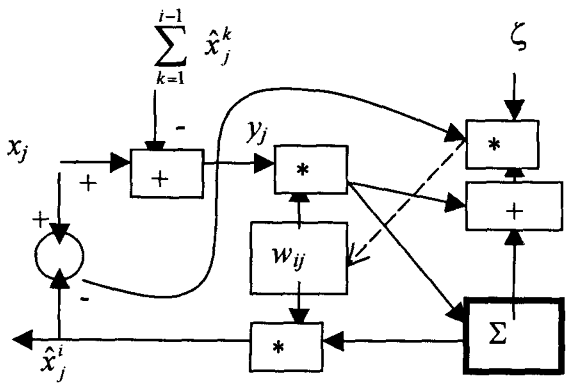

Figure 1: Single New PCA learning unit.

In Figure 1, the raw input data x_j is subtracted from the sum of the previous projected data on the previous principal components to obtain y as defined in the equation (2). The Σ box provides the inner product between vectors y and w_i. The result of the Σ box operation will, again, be summed with the previous multiplication of y_j and w_{ij} and its output will be multiplied with the learning rate ζ before updating to w_{ij} as described in equation (4). This single unit can be cascaded into n units to obtain a PCA learning vector and this learning vector can be cascaded to obtain many as parallel eigenvector extractors as needed for each application.

IV. APPLICATIONS

In this study, we used two gray scale images: Elaine and tank as shown in Figures (3a) and (4a). The purpose of this study is to evaluate how well our technique can extract the features of these images via principal components as opposed to the MATLAB technique and from the extracted features we can process for image compression.

Elaine image consists of 256x256 gray pixel and each pixel has 8-bit quantization. Tank is 512x512 pixel image with 8-bit gray scale/pixel.

We used input vector as row data with 64 pixel/vector to construct the training vector set. When the training vector set is available, the algorithm as shown in equation (4) is applied to extract the principal vector. Our study has shown that the maximum number of iterations required is 150 of learning repetitions and the first 20 component vectors are extracted.

Feature extraction

Feature extraction using PCA is a well known approach [1], and is based on the maximum statistics feature correlation.

The first 10-component vector extracted from Elaine image using our technique is projected onto the first 10-component from MATLAB (inner product) and its results are shown in Figure 2a.

As orthogonal characteristics between principal vectors, if the learning component vector and the component vector from MATLAB are the same order and identical (or close to identical), the expected inner product should be close to +/-1; otherwise, it should be close to zero.

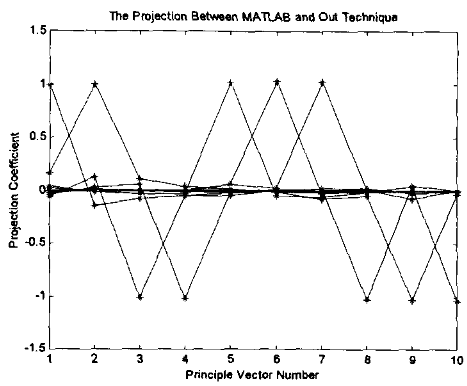

Figure (2a): The projection (inner product) result of 10-component vector extracted from Elaine image using our technique on 10-component vector from MATLAB.

The first 10-component vector extracted from tank image using MATLAB and our approach and the projection between principal vectors are shown in Figure 2b.
As shown in Figure 2a and 2b, there are ten values (+/-1) and the rest (70 values are close to zero) from which our technique can extract the feature vector as identical as that with the MATLAB technique.

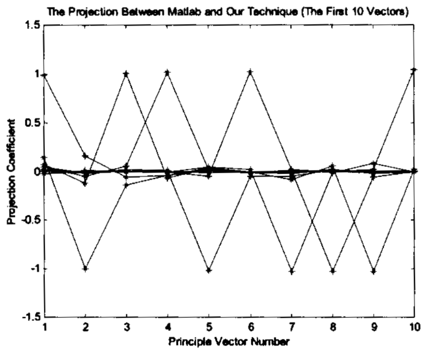

Figure 2b: The projection result of 10-component of our technique on 10-component of MATLAB (Tank image)

Compression

In this study, we extracted the first 20-component vector from full set of 64 component vectors. Since The full image is constructed from the first 20 component principal vectors extracted using MATLAB shown in Figures (3b and 4b) and using our approach (3c and 4c).

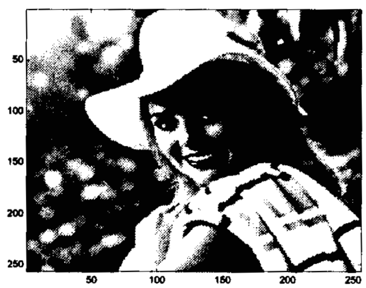

Figure 3a: original image (256x256)

Figure 3b: constructed image from 20-component using MATLAB

Figure 3c: constructed image from 20-component using our technique

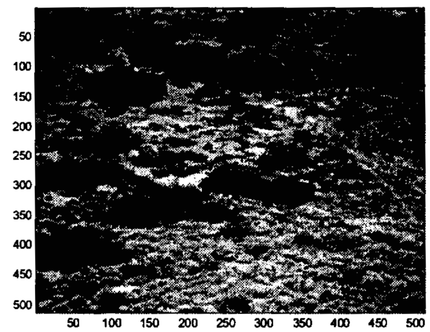

Figure 4a: original image (512x512)

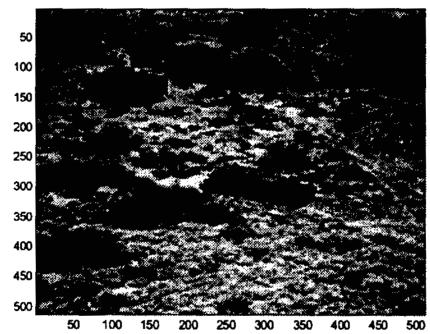

Figure 4b: constructed image from 20-component using MATLAB

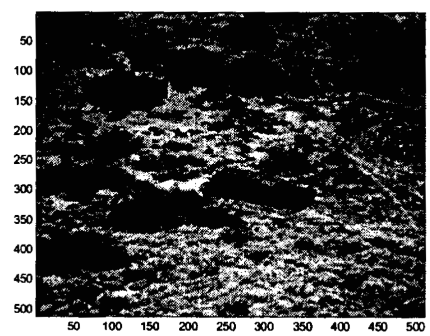

Figure 4c: constructed image from 20-component using our technique.

V. DISCUSSION

To demonstrate that our approach is more advanced as compared with the gradient descent technique, we have studied two given images (3a and 4a) using four techniques: GED, DOGED, GEDYN and DOGEDYN. The results are shown table I.

Table I: Comparison of number component to be extracted with complete convergence.

	GED	DOGED	GEDYN	DOGEDYN
Elaine	2	11	>20	>20
Tank	3	6	>20	>20

This study is based on 150 iterations for learning. The DOGED is better than GED in learning and hardware implementation as shown in Figure 1. When the dynamic initial learning rate is incorporated, the GEDYN and DOGEDYN are able to extract as many components as needed (or all component if needed).

VI. CONCLUSION

In this paper, we have demonstrated that our innovative DOGEDYN technique (which consists of *dominant element based gradient descent learning and dynamic initial learning rate*) requires much less hardware, simple system architecture, and reliable learning technique which can easily be implemented in VLSI hardware. Simplicity in hardware requirement is afforded because there are several identical building blocks as principal vector extractor and sit on the same chip to extract fully parallel spectral images. This will allow one to extract principal component vectors of hyperspectral data in a fully parallel fashion for real time classification based on the features or /and data compression.

VI. ACKNOWLEDGEMENTS

The research described herein was performed by the Jet Propulsion Laboratory, California Institute of Technology under contract with the National Aeronautics and Space Administration (NASA). The author would like to thank Drs Thomas Prince and Katherine Dumas for their support and Dr. T. Daud for his suggestions.

References:

1. A.K. Jain, R.P.W. Duin, and J. Mao "Statistical Pattern Recognition: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, NO. 1, January 2000.
2. A.K. Jain, *Fundamentals of Digital Image Processing.* Prentice Hall Information and System Sciences Series, 1989.
3. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, Newyork, Wisley, 1973.
4. P. Baldi and K. Hornik, "Learning in linear neural networks: A survey," *IEEE Trans. Neural Networks*, Vol. 6, pp. 837-857, 1995.
5. S. Haykin, *Neural Network-A Comprehensive Foundation.* New York: Macmillan, 1994.
6. C. Chatterjee et. al. "Algorithms for accelerated convergence of adaptive PCA." *IEEE Trans. Neural Networks*, Vol. 11, No. 2, pp. 338-355, March 2000.
7. Y. Chauvin, "Principal component analysis by gradient descent on a constrained linear Hebbian cell," in *Proc. Join Int. Conf. Neural Networks*, Vol. I, San Diego, CA pp. 373-380, 1989.
8. Z. Fu and E.M. Dowling, "Conjugate gradient Eigenstructure tracking for adaptive spectral estimation," *IEEE Trans. Signal Processing*, Vol. 43, pp. 1151-1160, 1995.
9. E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix." J. Math. Anal. Appl., vol. 106, pp. 69-84, 1985.
10. T.D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks*, Vol.2 , pp. 459-473, 1989.
11. L. Xu, "Least mean square error reconstruction principal for self-organizing neural-nets," Neural Networks, 6, pp. 627-648, 1993.
12. G. Vane, et. Al. "The airborne visible/infared imaging spectrometer (AVIRIS), Remote Sens. Environ. 44, pp. 127-143, 1993.
13. A. Howard, et al, "Intelligent Target Detection and Hyperspectra Imaging", Proc. 13th Int. Conf. Applied Geological remote Sensing, Vancouver, British Columbia, Canada, march, 1999.
14. T.A. Duong, "Dominant Element Based Gradient Descent Algorithm for Fast and Hardware Implementable PCA Technique," in preparation.

BIBLIOGRAPHY

Dr. Tuan Anh Duong received the Ph.D. degree from the UCI in 1995 in Electrical Engineering. Currently, he is a senior researcher of the Bio-Inspired Technologies and Systems Group at the NASA/Jet Propulsion Laboratory. He also holds an Adjunct Associate Professor position in the Electrical Engineering Department at University of Southern California. For his invention of on-chip processing to improve the cost of NASA missions he has received the 2002 NASA Space Act Award for the creative development of exceptional scientific and technical contribution for aerospace technology program of NASA. He has authored and co-authored over 51 technical reports and papers as well as two book chapters and he holds three patents and one patent pending. He is the recipient of over 11 NASA Technical Innovation Awards. His current research interests include real time on-chip learning neural networks, optimization dynamic scheduling, data compression using linear and non-linear techniques e.g., PCA and neural networks and VLSI hardware implementation as mixed signals system-on-a-chip approach using CMOS or SOI technologies.