International Journal of Innovative Computing, Information and Control
Volume **1**, Number **3**, September **2005**

ICIC International ©2005 ISSN 1349-4198
pp. **479–492**

QUANTIZED PRINCIPAL COMPONENT ANALYSIS WITH APPLICATIONS TO LOW-BANDWIDTH IMAGE COMPRESSION AND COMMUNICATION

Dave Wooden, Magnus Egerstedt

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
{ wooden, magnus }@ece.gatech.edu

Bijoy K. Ghosh

Department of Electrical and Systems Engineering
Washington University in St. Louis
St. Louis, MO 63130, USA
ghosh@netra.wustl.edu

Received November 2004; revised April 2005

Abstract. *In this paper we show how Principal Component Analysis can be mapped to a quantized domain in an optimal manner. In particular, given a low-bandwidth communication channel over which a given set of data is to be transmitted, we show how to best compress the data. Applications to image compression are described and examples are provided that support the practical soundness of the proposed method.*
Keywords: Principal component analysis, Quantization, Image compression

1. **Introduction.** Principal Component Analysis (PCA) is an algebraic tool for compressing large sets of statistical data in a structured manner. However, the reduction results in real-valued descriptions of the data. In this paper, we take the compression one step further by insisting on the use of only a finite number of bits for its representation. This is necessary in a number of applications where the data is transmitted over low-bandwidth communication channels. In particular, the inspiration for this work came from the need for multiple mobile robots to share visual information about their environment.

Assuming that the data $x_1, \ldots, x_N$ take on values in a d-dimensional space, one can identify the d principal directions, coinciding with the eigenvectors of the covariance matrix, given by

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)(x_i - m)^T, \tag{1}$$

where m is the mean value of the data.

If our goal is to compress the data set to a set of dimension $n < d$, we would pick the n dominant directions, i.e. the directions of maximum variation of the data. This results in an optimal (in the sense of least squared error) reduction of the dimension from d to n. For