IEEE Transactions on Geoscience and Remote Sensing

July 2012

Spherical Stochastic Neighbor Embedding of Hyperspectral Data

Lunga, D[1, 2] and Ersoy O[3]

1CSIR. Meraka Institute, Pretoria, South Africa

2School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-0501 USA

Abstract

In hyperspectral imagery, low-dimensional representations are sought in order to explain well the nonlinear characteristics that are hidden in high-dimensional spectral channels. While many algorithms have been proposed for dimension reduction and manifold learning in Euclidean spaces, very few attempts have focused on non-Euclidean spaces. Here, we propose a novel approach that embeds hyperspectral data, transformed into bilateral probability similarities, onto a nonlinear unit norm coordinate system. By seeking a unit l2-norm nonlinear manifold, we encode similarity representations onto a space in which important regularities in data are easily captured. In its general application, the technique addresses problems related to dimension reduction and visualization of hyperspectral images. Unlike methods such as multidimensional scaling and spherical embeddings, which are based on the notion of pairwise distance computations, our approach is based on a stochastic objective function of spherical coordinates. This allows the use of an Exit probability distribution to discover the nonlinear characteristics that are inherent in hyperspectral data. In addition, the method directly learns the probability distribution over neighboring pixel maps while computing for the optimal embedding coordinates. As part of evaluation, classification experiments were conducted on the manifold spaces for hyperspectral data acquired by multiple sensors at various spatial resolutions over different types of land cover. Various visualization and classification comparisons to five existing techniques demonstrated the strength of the proposed approach while its algorithmic nature is guaranteed to converge to meaningful factors underlying the data.