

# NONPARAMETRIC APPROACHES TO STYLE CONSISTENT CLASSIFICATION

By

Srinivas Andra

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Electrical, Computer, and Systems Engineering

The original of the complete thesis is on file  
in the Rensselaer Polytechnic Institute Library

Examining Committee:

George Nagy, Thesis Adviser

Kristin P. Bennett, Member

John W. Woods, Member

William A. Pearlman, Member

Rensselaer Polytechnic Institute  
Troy, New York

August 2006  
(For Graduation December 2006)

## ABSTRACT

Most pattern classifiers are trained on data from multiple sources, so that they can accurately classify data from any source. However, in many applications, it is necessary to classify groups of test patterns, with patterns in each group generated by the same source. The co-occurring patterns in a group are statistically dependent due to the commonality of source, which goes beyond the familiar linguistic and semantic contexts. The dependence between these patterns introduces *style context* within a group that can be exploited to improve the classification accuracy.

We propose a framework for style consistent classification with nonparametric methods for discrete style case, i.e., applications with a finite number of styles such that the sample size relative to each style is “sufficient”. Our work is motivated by previous work on style consistent classification with Gaussian mixture densities. The fewer assumptions on forms of underlying distributions and higher accuracies in most practical applications make nonparametric methods attractive candidates for extension to style consistent classification. Using this framework, we extend the Support Vector Machine (SVM) method – one of the most popular classification techniques proposed in the last decade – to style consistent classification. The nearest neighbor (NN) method, another nonparametric method with asymptotic classification error less than twice the optimal Bayes error, is also extended to style consistent classification using the same framework. Inspired by the geometry of class-and-style distributions in the feature space, we propose *frequency coding*, a novel nonparametric classification method, for single-label classification. Frequency coding lends itself naturally to style consistent classification within the proposed framework.

We conduct extensive simulations on pseudo-randomly generated data to gain insight into the benefits and the limitations of proposed methods. We further demonstrate that proposed methods yield optimal classification gains on machine-printed digits by reducing the error introduced by inter-source variations to essentially zero.