Geometric diffusion as a classifier

Alan Van Nevel
1900 N Knox Rd, M/S 6302, Naval Air Warfare Center, Weapons Division, China Lake, CA 93555, USA
E-mail: alan.vannevel@navy.mil

Abstract. Often, the problem of automatic target recognition can be reduced down to two separate but related problems, feature extraction and classifier design. The best classifier only works as well as the input data provided to the system. In this presentation, we will outline a new approach to classification known as geometric diffusion as proposed by Coifman et al, and demonstrate the power of this new metric for classification of imagery.

1. Introduction
In this note, we present some results on object recognition which are very good and move away from traditional machine learning approaches. We will first discuss geometric diffusion and its use as a classifier. The example problem of interest, vehicle recognition in laser radar (ladar) imagery will be described, followed by a discussion of the results.

2. Geometric Diffusion
Geometric diffusion [1,2] can be thought of as a dimensionality reduction technique which preserves the intrinsic geometry of the data under study. The approach is similar to other dimensionality reduction techniques such as Laplacian Eigenmaps [3], Diffusion wavelets [4], and Spectral Clustering [5]. The approach in geometric diffusion is to develop a data dependent kernel which utilizes the spectral properties of the kernel to determine the underlying geometry of the data. This technique offers several advantages for use in dimensionality reduction, feature extraction and manifold applications. The main advantage is an extrinsic distance measure is used to recover the intrinsic data geometry, without the use of an a priori model for the data.

2.1. Basic Theory
The goal is to embed the data into a lower dimensional space, using a diffusion operator constructed from the data dependent heat kernel. The kernel \( K \), needs to meet the following requirements: symmetric, positive semi-definite, and positivity preserving. A typical kernel that can be used to define distances between data points is

\[
K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\epsilon} \right)
\]

\( ^* \) Cleared for public release, distribution unlimited
where $x_i$, $x_j$ represent the data under study. Other kernel functions can be defined as needed as long as they meet the conditions. The diffusion operator then is written as

$$A = \left[ \frac{K(x_i, x_j)}{\sum_j K(x_i, x_j)} \right] = D^{-1}K$$

(2)

It is this operator $A$, that is critical for the classification task. Given a partially labeled data set $X = \{x_j\}, \ i = 1 \ldots N$, where $N$ is the number of data points, consider a label function $\chi_j(x)$ that equals 1 when $x$ is a member of class $j$, and is zero otherwise. The procedure then is to smooth $\chi_j(x)$ using powers of $A$

$$\chi'_i = A^t \chi_i$$

(3)

where the power $t$ is chosen typically by cross validation.

In this approach, not all of the data needs to be assigned a label. As long as each class has one exemplar provided, the smoothing operation will attempt to assign labels to each data point based on the underlying geometry of the data [6].

3. Ladar problem and results

Ladar sensors have the capability to capture 3-d information, which should provide for increased object recognition capabilities. Using synthetic ladar imagery, a five class recognition problem was developed, using 5 different military vehicles. Two different data sets were constructed, one where the depression angle varied from 5 to 52.5 degrees, at a fixed aspect, and the second had a fixed depression angle, and the aspect varied from 45 to 135 degrees. Each class had 20 examples. The imagery consisted of 32x32 pixel images. The images were lexicographically ordered by column, and the diffusion operator $A$ constructed.

The performance was assessed as a function of the number of labelled exemplars provided per class, and averaged over 1000 runs, randomly selecting the images to be assigned a label. The remaining data was then classified by smoothing the label function with the diffusion operator. In each data set, as the number of labelled exemplars approached the maximum of 19 (equivalent to leave one out validation), the classification performance became 100%, and remained above 95% using as few as 7 exemplars. Using only one exemplar, performance was above 70% for the depression data.

4. Conclusions

The geometric diffusion ideas provide a powerful new approach to clustering high dimensional data into a lower dimensional space while preserving the underlying intrinsic geometry. The approach permits improved classification performance for difficult image recognition tasks. The major drawback to this approach is the limited ability to add new data to the kernel. For limited recognition tasks, using small data sets, one can simply recompute the kernel, which quickly becomes unfeasible when dealing with very large high dimensional data.

References

[1] Coifman R and Lafon S 2006 Appl Comp Harm Anal 21 pp 5-30
[2] Lafon S “Diffusion Maps and Geometric Harmonics”, Ph.D. Dissertation, Yale Univ, 2004
[3] Belkin M and Niyogi P 2003 Neural Computation 15(6) pp. 1373-96
[4] Coifman R and Maggioni M 2006 "Diffusion Wavelets", Appl Comm Harm Anal 11
[5] Chung F, “Spectral Graph Theory”, CBMS-AMS, 92, 1997
[6] Szlam A, Maggioni M and Coifman R 2008 J.I of Machine Learning Research 9 pp1711-38