Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Image and Vision Computing 25 (2007) 365–373

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Image distance functions for manifold learning

Richard Souvenir, Robert Pless *

Department of Computer Science and Engineering, Washington University, One Brookings Drive, Campus Box 1045, St Louis, MO 63130, USA

Received 21 October 2004; received in revised form 4 May 2005; accepted 5 January 2006

Abstract

Many natural image sets are samples of a low-dimensional manifold in the space of all possible images. When the image data set is not a linear combination of a small number of basis images, linear dimensionality reduction techniques such as PCA and ICA fail and non-linear dimensionality reduction techniques are required to automatically determine the intrinsic structure of the image set. Recent techniques such as ISOMAP and LLE provide a mapping between the images and a low-dimensional parameterization of the images. This paper specializes general manifold learning by considering a small set of image distance measures that correspond to key transformation groups observed in natural images. This results in more meaningful embeddings for a variety of applications.

PACS: 02.60.Ed; 87.80Pa

Keywords: Isomap; Manifolds; Non-parametric registration

1. Introduction

Faster computing power and cheap large-scale memory have led to a surge in research in the machine learning community on the topic of dimensionality reduction, which finds structure in a large set of points embedded in a very high-dimensional space. Many problems in computer vision can be cast in this framework, as each image can be considered to be a point in a space with one dimension for each pixel. When an image data set is generated by varying just a few parameters, such as a combination of pose, lighting, or camera viewpoints, then this set can be considered to be sampling a continuous manifold of the space of all possible images. Given a set of images, understanding this manifold and automatically parameterizing each image by its place on this manifold has emerged as an important tool in the model-free interpretation of image data.

Especially for the analysis of the variation in images of a single object, this approach was long foretold:

> In a very large part of morphology, our essential task lies in the comparison of related forms rather in the precise definition of each; and the deformation of a complicated figure may be a phenomenon easy of comprehension, though the figure itself may have to be left unanalyzed and undefined. —D'Arcy Thompson [11]

Algorithms for inferring properties of image manifolds by comparing related images have been codified in a family of computational techniques exemplified by Isomap [10] and locally linear embedding (LLE) [8]. These techniques extend a sparse set of local relationships between similar images to a global low-dimensional parameterization of all images. This work uses Isomap as an exemplar of this class of non-linear dimensionality reduction tools, and the results will directly apply to other methods including semidefinite embedding [14], and could be extended to LLE, alignment of local representations (LLC) [9], and Hessian Eigenmaps [4].

The main contribution of this paper is to explore the application of Isomap to video imagery, and to guide the process of specializing Isomap for particular problem domains. Several earlier papers have visualized the parameterization of image sets and observe that it highlights perceptually relevant features. Here, we emphasize that the parameterization produced by Isomap is a function of the input data set and the image distance metric.

A formal theory of the statistics of natural images and natural image variations—pattern theory—gives tools for defining relevant image distance metrics. We postulate that for natural image data sets, a small number of distance metrics

* Corresponding author. Tel.: +1 314 935 7546; fax: +1 314 935 7302.
E-mail address: pless@cse.wustl.edu (R. Pless).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.01.016

are useful for many important applications. This paper proposes a set of distance measures that correspond to the most common causes of transformation in image sets and gives examples of how these significantly improve performance on a variety of application domains, including the de-noising of cardiac MR imagery.

2. Differential structure in dimensionality reduction

Dimensionality reduction is an important tool in image data analysis, because images are large, and when treated as a vector of pixel intensity values, lie in a very high-dimensional space. Here, we give a brief introduction to Isomap as one tool for non-linear dimensionality reduction. This is explicitly compared with linear dimensionality reduction as typified by principal components analysis (PCA). We argue that PCA is poorly suited to the analysis of many natural image sets, especially those which include motion. We then consider the structure of both Isomap and PCA embeddings.

2.1. Background of Isomap

Given an input set $\mathcal{I}$, which is a finite subset of $\mathcal{R}^D$, (where D is the number of pixels in an image), the dimensionality reduction techniques of Isomap and LLE produce a mapping function $f:\mathcal{I} \rightarrow \mathcal{R}^d$. Very briefly, Isomap begins by computing the distance between all pairs of images (using the square root of the sum of the squared pixel errors, which is the L_2 norm distance if the images are considered points in $\mathcal{R}^D$). Then, a graph is defined with each image as a node and undirected edges connecting each image to its k-closest neighbors (usually choosing k between 5 and 10). A complete pair-wise distance matrix is calculated by solving for the all-pairs shortest paths in this sparse graph. Finally, this complete distance matrix is embedded into $\mathcal{R}^d$, by solving an eigenvalue problem using a technique called multidimensional scaling (MDS) [3]. d is the dimension of the low-dimensional embedding and can be chosen as desired, but, ideally, is the number of degrees of freedom in the image set. LLE is a method with similar aims that creates a mapping that preserves linear relationships between nearby points. The original papers for Isomap [10] and LLE [8] have pointers to online, free implementations of the algorithm, a tradition which has been continued for the successors of these algorithms, including Hessian Eigenmaps [4], Laplacian Eigenmaps [1], and semidefinite embedding [14].

Isomap has several very important limitations. First, the Isomap algorithm defines a mapping from the original image set to $\mathcal{R}^d$. That is, Isomap computes a mapping $f:\mathcal{I} \rightarrow \mathcal{R}^d$ and not, as might be more convenient, $f:\mathcal{R}^D \rightarrow \mathcal{R}^d$. This means that once the Isomap embedding of an image set $\mathcal{I}$ is computed, for $I' \notin \mathcal{I}$ the value of $f(I')$ is not well defined. Additionally, the inverse mapping is also problematic. For a point $x \in \mathcal{R}^d$, if x is not in the set of points defined by $f(\mathcal{I})$, then $f^{-1}(x)$ is also not well defined. Although approaches have been proposed to compute these 'out of sample' projections [2], this remains, both theoretically and practically, a challenge for Isomap and other dimensionality reduction techniques.

2.1.1. Comparison to PCA

It is instructive to view PCA in the same light. Given an input set of images $\mathcal{I}$ (still a finite subset of $\mathcal{R}^D$), principal component analysis computes a function f which projects each image onto a set of basis images. The image set, $\mathcal{I}$, is used to derive a set of orthonormal basis images $\vec{b}_1,\vec{b}_2,\ldots,\vec{b}_d$, and then the function f which maps an image x in $\mathcal{R}^D$ to a set of coefficients in $\mathcal{R}^d$ is:

$$f(x) = (x^\top \vec{b}_1, x^\top \vec{b}_2, \ldots, x^\mathrm{T}\vec{b}_d) = (c_1, c_2, \ldots, c_d)$$

Therefore, although the basis images are defined based upon an Eigen-analysis of the image data set $\mathcal{I}$, the function f is defined for all possible images of D pixels:

$$f_{\mathrm{PCA}}:\mathcal{R}^D \rightarrow \mathcal{R}^d$$

In addition to being more computationally efficient, the projection function f of PCA remains well defined for images that are not present in the original set $\mathcal{I}$. Also, the inverse function is defined as well, so that any point in the coefficient space can be mapped to a specific image by a linear combination of basis images:

$$f^{-1}_{\mathrm{PCA}}(c_1, c_2, \ldots, c_d) = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_d\vec{b}_d \quad (1)$$

Differential changes to the coefficients correspond to changes in weights of the linear basis functions. Consider an image x with corresponding coefficients $(c_1, c_2, \ldots, c_d)$. The partial derivative of the inverse mapping function (Eq. (1)) describes how the image varies when changing the c_i coefficient:

$$\frac{\partial}{\partial c_i} f^{-1}_{\mathrm{PCA}}(c_1, c_2, \ldots, c_d) = \vec{b}_i$$

Equivalently, moving through the coefficient space can be interpreted as an operator: changing coefficient c_i by ϵ changes the image x by the addition of the b_i basis image: $x' = x + \epsilon\vec{b}_i$.

However, this is not usually the type of image change that underlies natural image variations. Natural changes to images, for example, those due to variation in pose or shape deformations, are very poorly approximated by changes in linear basis functions. Fig. 1 shows the PCA decomposition of an icon moving smoothly from left to right. Despite the fact that this image set has only one degree of freedom, it takes many principal components to reconstruct any of the original images effectively. This leads to the question: what local variations dominate the relationships between similar images in natural settings?

2.2. Differential structure in image manifolds

Non-linear dimensionality reduction, despite its drawbacks, has been successful at finding natural parameterizations, or 'perceptual organizations' [10], of a variety of different image sets, including pose estimates in rigid body motions [7],