Principal Components Analysis of Scalar, Vector, and Mesh Vertex Data

Release 1.00

Michael Bowers[1], Laurent Younes[1]

September 17, 2012

[1]Johns Hopkins University, Baltimore, MD

Abstract

This document describes a contribution to the Insight Toolkit intended to support the analysis of the principal components of data sets, optionally point data associated with the vertices of a mesh.

This paper is accompanied with the source code, input data, parameters and output data that we used for validating the implementation described in this paper. This adheres to the fundamental principle that scientific publications must facilitate **reproducibility** of the reported results.

Contents

1 Introduction

A major goal of CA is to create algorithmic tools to aid basic and clinical neuroscientists in the analysis of anatomical structures at different scales. The main difficulty is the complexity of anatomical substructures and the large variation between individuals. Our group championed the idea that anatomical structures can be represented as a collection of coordinate systems: landmark points (0D), curves (1D), surfaces (2D), and sub-volumes (3D). Anatomical variability can be characterized by diffeomorphic transformations of these coordinate systems. Thus anatomies are represented as deformable templates, with the space of anatomical images being the set generated by the group of diffeomorphic transformations acting on the template with associated probability laws, which describe how they vary. The transformations are detailed so that a large family of shapes may be generated with the precise topology of the template maintained. The diffeomorphic transformations give rise to vector field correspondences which can be expanded as a complete orthonormal basis $V(x) = \Sigma_{i=1}^{N} V_i \phi_i(x)$ where x lies on the template, $\phi_i(x)$ are the shape functions and V_i are independent Gaussian random variables with fixed means and covariances. The maps are represented via the N-vectors of coefficients $(V_1, \ldots, V_N)$. Gaussian hypothesis testing may then be applied to these coefficients.

This paper describes a class itk::VectorFieldPCA, an implementation of standard PCA algorithms for use on scalar or vector data sets. Kernel PCA is implemented in this class as well, where the data sets are scalar or vector valued functions assigned at each of the points in a `PointSet`. A Gaussian Distance Kernel class is provided with the PCA class.

This contribution is part of a shape analysis software pipeline created at Johns Hopkins. PCA will be used to develop a set of basis vectors for use with Gaussian Random Field analysis. The output of PCA will be analyzed for significance with various statistical methods such as t-tests built upon the built-in statistical capabilities of ITK.

2 Template Parameters

This class is templated over the types of the vector valued functions, the output data types, and optionally the point set type. An optional kernel function type parameter defaults to `itk::KernelFunction`.

3 Inputs

Input to the class consists of sets of scalar or vector valued data. The user can set an optional kernel function to invoke Kernel PCA, and a `itk::PointSet` for kernel operations.

4 Outputs

- Average of the vector/scalar measurements over all the data sets
- The eigenvalues of the Principal Components Analysis
- The set of basis vectors of the data set
- The projection of the basis onto the vector field set, or any vector set specified in a call to `Projection()`.

5 How to Build

This contribution includes

- Source code for the PCA function
- Tests for the code
- Test data and output

5.1 Building Executables and Tests

In order to build the whole, it is enough to configure the directory with CMake. As usual, an out-of-source build is the recommended method.

In a Linux environment it should be enough to do the following:

- `ccmake SOURCE_DIRECTORY`
- `make`
- `ctest`

Where SOURCE_DIRECTORY is the directory where you have expanded the source code that accompanies this paper.

You will be required to provide the directory where you built or installed ITK.

- `ITK_DIR`

This will configure the project, build the executables, and run the tests and examples.

5.2 Building this Report

In order to build this report you can do

- `ccmake SOURCE_DIRECTORY`
- Turn ON the CMake variables
  - `GENERATE_REPORTS`
- `make`

This should produce a PDF file in the binary directory, under the subdirectory `Documents/Report001`.

6 PCA Test Code

The source code used to test this function provides a good example of its use. The scalar PCA test operates on a set of scalar defined functions in the standard way. The vector kernel test uses the Gaussian distance kernel that comes with the class, so a vtkPolyData mesh file is input, as well as a list of text files containing vector values functions defined at every vertex in the mesh.

After the PCA calculation, the outputs are available and the test programs write them out as text data.

The source code presented in this section can be found in the `Testing` directory under the filenames

- `itkVectorKernelPCATest.cxx`

```
USAGE: VectorKernelPCA <pcaCount> <kernelSigma> <vtkMeshFile> <outputName> <vectorFieldSe
    pcaCount : number of principal components to calculate
    kernelSigma : KernelSigma (width of Kernel, usually about 6.25)
```

6.1 Results

Figure 1 shows the averages for the vector valued inputs to the test program

```
VectorKernelPCA 5 6.75
    PCATestSurface.vtk vectorPcaOutput
    PCATestSurface_alpha0_01.vtk PCATestSurface_alpha0_02.vtk PCATestSurface_alpha0_03.vtk PCA
    PCATestSurface_alpha0_06.vtk PCATestSurface_alpha0_07.vtk PCATestSurface_alpha0_08.vtk PCA
    PCATestSurface_alpha0_10.vtk PCATestSurface_alpha0_11.vtk PCATestSurface_alpha0_12.vtk PCA
    PCATestSurface_alpha0_15.vtk PCATestSurface_alpha0_16.vtk PCATestSurface_alpha0_17.vtk PCA
    PCATestSurface_alpha0_20.vtk PCATestSurface_alpha0_21.vtk PCATestSurface_alpha0_22.vtk PCA
    PCATestSurface_alpha0_25.vtk PCATestSurface_alpha0_26.vtk PCATestSurface_alpha0_27.vtk PCA
    PCATestSurface_alpha0_30.vtk PCATestSurface_alpha0_31.vtk PCATestSurface_alpha0_32.vtk PCA
    PCATestSurface_alpha0_35.vtk PCATestSurface_alpha0_36.vtk PCATestSurface_alpha0_37.vtk PCA
```

The PCATestSurface.vtk input file template altas of a caudate surface. The alpha files contain the initial momenta vectors at each vertex determined during the calculation of a deformation between a template and a set of target caudate surfaces. The averages over all the target momenta are displayed in CAWorks, a JHU Center for Imaging Science Paraview-based application.

7 Acknowledgements

Funding for development provided by NIH grants (R01-EB008171-01A1 and P41-RR015241).

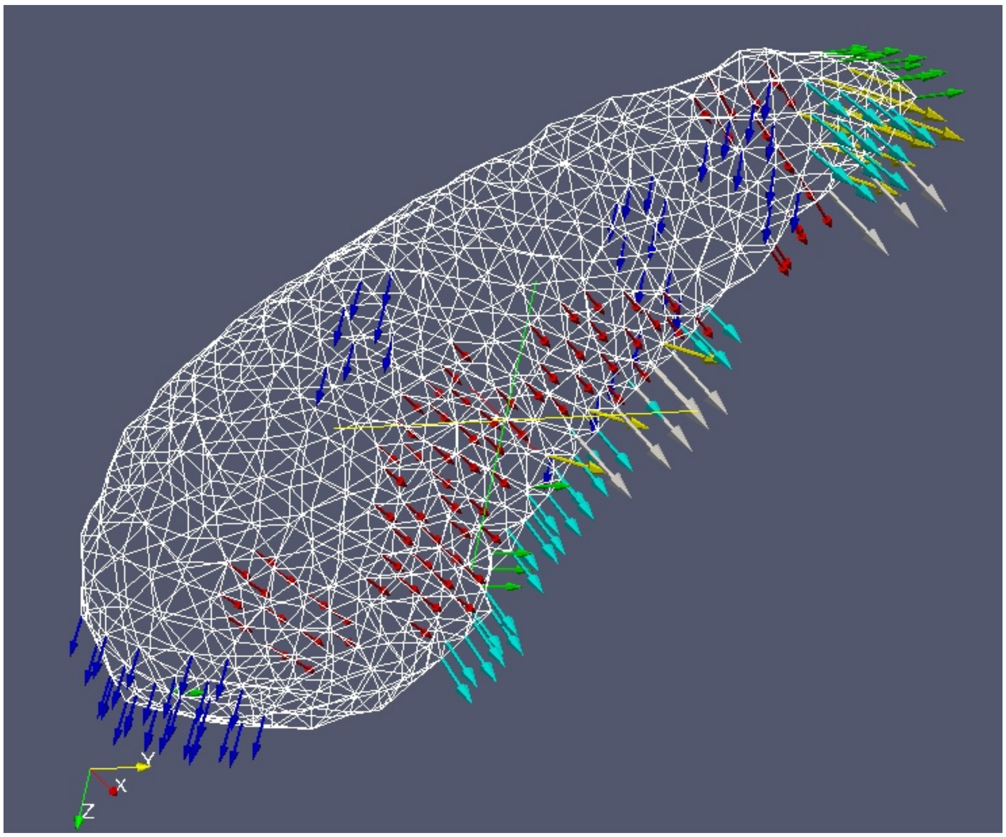

Figure 1: First PC of momenta on a triangulated caudate surface.