Principal Component Analysis

Implementation in CGAL, and application to surface normal estimation and dimension detection

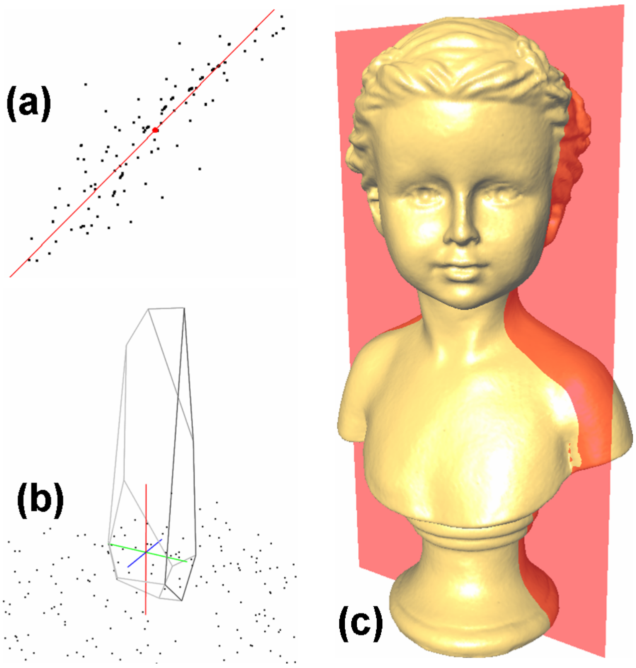
GEOMETRICA, INRIA Sophia-Antipolis

Principal component analysis (PCA) is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones.

PCA for Geometric Computing. PCA plays a key role for a class of geometric computing algorithms, in particular for *linear least squares fitting*. Given a point set in arbitrary dimension, linear least squares fitting amounts to find the linear sub-space which minimizes the sum of squared distances from the points to their projection onto this linear sub-space (Figure a). This problem is equivalent to search for the linear sub-space which maximizes the variance of projected points, the latter being obtained by eigen decomposition of the covariance matrix.

The CGAL library (`http://www.cgal.org`) already contains a set of functions for linear least squares fitting point sets in 2D, point sets in 3D as well as triangle sets in 3D (Figure c). Notice that for a triangle set this problem is equivalent to the one of fitting a linear sub-space to a point set, except that the covariance tensor is derived from a continuous integral over the triangles instead of a discrete sum over the points. The main interest is to get independence to the density of triangles for a surface triangle mesh.

Goals. The first goal is to extend the current CGAL PCA package to compute covariance tensors of other simple geometric primitives such as segments, triangles, disks in 2D; segments, tetrahedra, balls, ellipsoids and polyhedra in 3D; and point sets in arbitrary dimension. Time permitting, we will further enrich the PCA package with computations of other geometric properties such as inertia tensors (derived from covariance tensors) and higher order moment tensors. The second objective is to turn the enriched PCA package at work.



Applications. The first application is normal estimation from point sets coming from laser scanners (Figure b) where the normal direction estimate (in red) is obtained by computing the covariance tensor of a polyhedron taken as the 3D Voronoi cell. We will compare such normal estimator with the state-of-the-art, and check its relevance for surface reconstruction. The second application is shape dimension detection. It is planned during the internship to scan some physical shapes using a laser scanner available at Genova, Italy (our partner in a European network of excellence).

Required Skills and Qualifications. Required skills include a knowledge of linear algebra, taste for geometry, and hands on C++ programming experience. Generic programming and STL experience a plus. Required qualification is a BS/MS degree in Computer Science or equivalent.

Contact. Pierre Alliez, researcher in geometric computing and processing. Tel: +33 492 387 677. Feel free to ask any question you may have about this internship by e-mail: pierre.alliez@sophia.inria.fr).