Title:
A Category-Major Interoperable Data Format for Optic Morphometry

Authors:
John K. Johnstone, jkj@uab.edu, Computer Science and Ophthalmology, UAB

Keywords:
Data management, Morphometry, Optic Nerve Head, Data format, Category-Major, PLY format

DOI: 10.14733/cadconfP.2024.121-125

Introduction:

A good data format is an important catalyst to computation and to data sharing. A good data format should be clear, concise, well documented (optimally as part of the format itself), should reflect the desired internal structure of the data, and should ideally adopt a familiar and widely accepted standard format, to encourage simple adoption by the research community. In this paper, we develop a new data format for ONH datasets with these characteristics, adapting the widely accepted Stanford PLY format.

An ONH dataset is a special form of point cloud, gathered from images of the optic nerve head. The optic nerve head is first imaged into a volume using spectral domain optical coherence tomography (OCT); next, this volume is sliced into images by planes rotating around an axis through the middle of the optic nerve head; and finally, the ONH structures of interest are segmented in each of these images (typically by hand since this segmentation task is subtle and challenging). This defines a point cloud organized both into categories (the structures of interest in the optic nerve head) and slices (the planar slices through a category), which we call an *ONH dataset*. For ONH datasets, it makes sense for the planar slices to rotate around an axis through the middle of the optic nerve head, rather than to lie in parallel planes as is typical in contour datasets, to focus the gathering of data to the cylinder of the optic nerve head. Figure 1 shows an ONH dataset and its organization into categories and slices.

Since the development of glaucomatous optic neuropathy is associated with structural changes in the optic nerve head, such as remodeling of the lamina cribrosa, the study of the morphology and morphometry of the optic nerve head is an active field of research in ophthalmology, requiring reconstruction and analysis of the optic nerve head using techniques from shape modeling and CAD. The data for this research is stored in ONH datasets, and this paper considers the storage and management of this data.

The rest of the paper is structured as follows. The next section discusses the structures of interest in the optic nerve head, and the associated morphometric measurements. After contrasting slice-major and category-major storage of an ONH dataset, we discuss the existing format for ONH datasets, the PLY format, and the new format for ONH datasets. We end with conclusions and future directions.

The structures of interest:

The four most important structures of interest in the optic nerve head are the inner limiting membrane (ILM), anterior lamina (AL), Bruch's membrane (BM), and anterior sclera (AS). In the language of ONH datasets, these are called categories. Reference structures for morphometry are defined from Bruch's membrane or anterior sclera (a reference plane and cylinder either from the opening of Bruch's membrane

[3] or from the anterior sclera [2]). Key morphometric measurements are then defined from the inner limiting membrane and anterior lamina, using these reference structures: such as cup depth (signed distance of ILM from the reference plane), laminar cup depth (distance of AL from the reference plane), cup volume (volume within the reference cylinder from the reference plane to ILM on the laminar side), and laminar cup volume (volume within the reference cylinder from the reference plane to AL). The four main structures (ILM, AL, BM, AS) are visible in all images (OCT or histology), and other structures of lesser interest may also be included, such as the anterior scleral canal opening or nerve fiber layer. Histological datasets typically contain more structures, such as posterior lamina, posterior sclera, neural boundary, and external boundary of pial sheath. Therefore, a data format must be flexible enough to allow a variable number of categories.

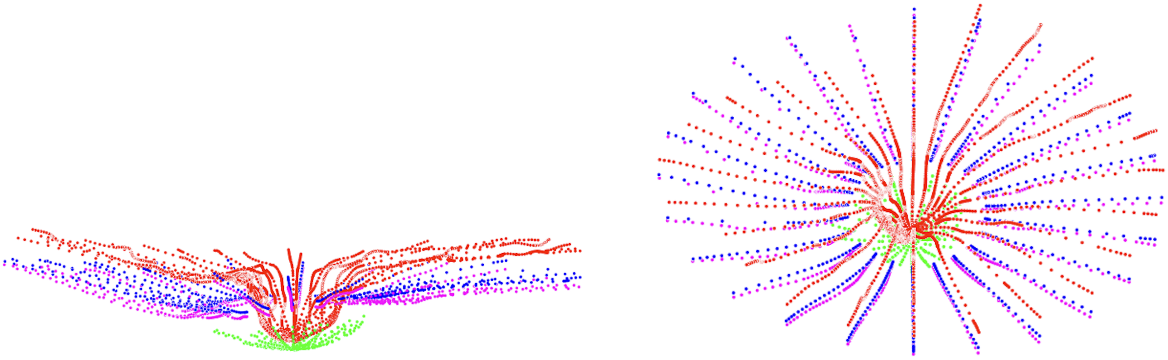

Fig. 1: An ONH dataset with four categories. Left: The categories: ILM (red), BM (blue), AS (magenta), and AL (green) Right: A different viewpoint of this dataset, showing its organization into radial slices.

Slice-major vs category-major storage:

Since an ONH dataset consists of categories and slices, the storage of its point cloud may be organized either by category or by slice. Since this is analogous to row-major vs column-major storage of a matrix, we call these two storage orders category-major and slice-major. In *category-major* storage, the category moves more slowly (just as in row-major storage, the row moves more slowly) while in *slice-major* storage, the slice moves more slowly.

The existing format for ONH datasets is slice-major. This mirrors the acquisition of the ONH dataset from OCT (or histology) through image segmentation of each planar slice. That is, the point cloud is sorted by slice, then by category within each slice. The existing format is also bespoke, in a local format that is not standard for point clouds.

A slice-major data format for ONH datasets is awkward, since computation with the optic nerve head (e.g., the extraction of a reference plane and ellipse from the opening of Bruch's membrane, or surface reconstruction of ILM for computation of cup depth) is more category-centric. Therefore, the new data format pivots to a category-major representation. The new format also adopts a standard data format for point clouds and polygon meshes, Stanford's PLY format [4], that is widely used in computer graphics and shape modeling, with existing tools such as readers [1] and familiary within the research community. Another nice feature of the PLY format is that it allows for natural documentation of the semantics of its data entries within the dataset itself, in a lightweight fashion. This adds to the clarity of the data format.

In short, we shift from a bespoke acquisition-based slice-major format to a standardized computation-

Proceedings of CAD'24, Eger, Hungary, July 8-10, 2024, 121-125

based category-major format. The next sections describe the two formats, existing and new, in more detail.

The present format for ONH datasets:

In this section, we describe the existing format for ONH datasets. An ONH dataset is a point cloud, and each point is a sample from the surface defined by a certain ONH category, which lies on a certain slice. The existing format is organized as follows. Each line of the file represents a point and has 6 entries. Here is a typical line:

```
5011.16 4816.23 1096.28 30 4 0
```

The first three entries of each line are the Cartesian coordinates of the point; each is a floating point scalar.[1] The fourth entry of each line is the category code of the point, indicating the category that the point lies in. The fifth entry of each line is the 0-based slice index of the point. Although the increment between consecutive slice indices is consistent, the typical increment is 4, not 1, and could be a float like 4.5. The sixth entry of each line is a boolean mark, associated with a special property of the point during segmentation. It is ignored by all morphometric computations.

The format is slice-major: all points of the first slice are recorded, then all points of the second slice, and so on. Within a slice, the points are sorted by category code: for example, the first points of the dataset are the points of the first slice from the category with the lowest category code in the dataset, sorted along the slice.

It turns out that the fifth and sixth entries (slice index and boolean mark) are both vestigial, since they are not used in any of the standard morphometric computations and do not influence any geometry. Like the organization of the dataset into slices, they are another aspect of the present format that reflects the segmentation process. Therefore, these components of the dataset are stripped from the new format.

PLY format:

The new data format is an adaptation of the PLY format, a format developed by the Stanford Computer Graphics Laboratory [4], which was originally designed for polygon meshes but, anticipating unknown future needs, was made flexible enough that it could represent point clouds and essentially any geometric data. PLY is a standard in shape modeling and computer graphics, widely adopted because of its flexibility, its rigour, and its ability to document a format clearly. This has also led to software for handling PLY files, such as Sharp's happly [1]. Another nice feature is that data can be stored in plain text or in binary. This section gives a concise description of PLY format; more details are available through various sources including Turk's official documentation [5, 6]. The next section discusses how we have adapted PLY format for ONH datasets.

A PLY file is organized into elements. A PLY file begins with a header, which describes each element in the data, including its name, a count of the number of these elements in the object, and a list of the various properties associated with the element, with their types. Arbitrary comments are also allowed in the header. For example, the famous Stanford bunny [4] is a triangle mesh whose PLY file has two elements (vertex and face), with the vertex element having 5 properties (Cartesian coordinates, confidence, and intensity, each a 32-bit float) and the face element having one property (an integer list, of the indices of the vertices of this face). The header is followed by the data itself, in this case 35,947 vertices and 69,451 faces.

The proposed format for ONH datasets:

[1] The coordinate frame of these points is not important since, in an early stage of computation, the dataset will be shifted to another coordinate frame, defined by the reference structures.

The proposed data format for ONH datasets adapts the PLY format. Like the PLY polygon mesh, it defines a point cloud using the vertex element. Like the PLY polygon mesh, it imposes a structure on the point cloud using other elements, but instead of using a face element to define polygons, it uses a section element to define sections (a section is the point cloud of a certain category in a certain slice), and a category element to define categories. By leveraging the added structure of a category-major point cloud, the definition of sections and categories is streamlined, reducing to a single integer per section and a single integer per category.

We now define the three elements of an ONH dataset more precisely (Table 1). It may be useful to follow along with the example of an ONH dataset in Figure 2; this dataset contains 6 categories, each with 12 sections, and 3821 points in the entire point cloud.

Element	Properties	Type
vertex	x, y, z	double, double, double
section	start	integer list
category	code	integer list

Table 1: The elements and properties of an ONH dataset in PLY format

```
ply
format ascii 1.0
element category 1
property list uchar int code
element section 6
property list uchar int start
element vertex 3821
property double x
property double y
property double z
end_header
6 30 31 45 50 51 120
12 0 34 65 93 124 164 207 247 293 348 391 428
12 459 508 542 567 584 603 640 677 723 754 794 822
12 859 874 907 923 940 956 967 979 991 1000 1010 1020
12 1034 1054 1079 1101 1126 1151 1173 1190 1215 1233 1246 1271
12 1286 1308 1333 1358 1383 1408 1433 1455 1477 1496 1521 1549
12 1574 1863 2089 2306 2454 2623 2762 2916 3055 3254 3393 3586
5216.010 4784.190 1103.990
```

Fig. 2: The header and first few entries of an ONH dataset in the proposed PLY format

The vertex element of an ONH dataset represents the entire point cloud of samples (all categories, all slices). It has three properties x, y, and z, representing the Cartesian coordinates. The point cloud is assumed to be in category-major order: by category, then by slice within the category. This element is analogous to the vertex element of a polygon mesh.

The section element of an ONH dataset represents the sections of a single category. Each section may be described by a single integer, the 0-based index of the first point of the section. This leverages the category-major structure of the point cloud. Only the start index is necessary since the end index is implied by the start index of the next section, and the points of a section are contiguous. If there is no next section, then the end index of the present section is implied by the size of the point cloud, which is specified in the vertex header element.

The section element has one property (start), which is a list of the start indices of the sections of the category in question. Therefore, each entry of the body associated with the section element will be a list of integers of length n_s, where n_s is the number of rotational slices through the ONH volume (often 12). For example, in the example of Figure 2, the third section entry lists the 12 sections of the anterior lamina (45), and the first of the 12 sections runs from the point with index 859 to the point with index 873.

The category element represents the order of the categories in the point cloud. It has one property (code), which is a list of the category codes.

Note the terseness of the format, due to the hierarchical tree structure imposed on the category-major ordering of the point cloud: a single integer is sufficient to encode each section and each category. Additional documentation of the dataset may be added as comments in the header.

The new format has a lower space complexity than the existing format: nPt*3 + nSec + nCat, rather than nPt*6. For example, if a dataset contains 10,000 points, 6 categories, and 72 sections (12 per category), then the existing dataset would require 60,000 entries, while the PLY dataset would require 30,078 entries. More importantly, the point cloud has a category-major structure, and the dataset's contents are well documented.

Conclusions:

We have developed a new PLY-based category-major data format for optic nerve head data. The proposed format is promising for data management and sharing of funded research involving optic morphometry, since its foundation is a widely used standard already accepted by the research community.

References:

[1] hapPLY: https://github.com/nmwsharp/happly.

[2] Johnstone, J.; Fazio, M.; Rojananuangnit, K.; Smith, B.; Clark, M.; Downs, J.C.; Owsley, C.; Girard, M.; Mari, J.-M.; Girkin, C.: Variation of the Axial Location of Bruch's Membrane Opening with Age, Choroidal Thickness and Race. Investigative Ophthalmology and Visual Science, 55:3, 2014, 2004–2009. https://doi.org/10.1167/iovs.13-12937

[3] Johnstone, J.; Rhodes, L.; Fazio, M.; Smith, B.; Wang, L.; Downs, J.C.; Owsley, C.; Girkin, C.: Measuring Mean Cup Depth in the Optic Nerve Head. CAD 2015, 2015, 216–221. https://doi.org/10.14733/cadconfP.2015.216-221

[4] The Stanford 3D Scanning Repository: https://graphics.stanford.edu/data/3Dscanrep

[5] Turk, G.: The PLY Polygon File Format.
Archived at https://gamma.cs.unc.edu/POWERPLANT/papers/ply.pdf.

[6] Wikipedia article on PLY file format. https://en.wikipedia.org/wiki/PLY_(file_format)