*Bioinformatics*, 36(14), 2020, 4189–4190
doi: 10.1093/bioinformatics/btaa521
Advance Access Publication Date: 16 May 2020
Applications Note



Genome analysis

# Stereo3D: using stereo images to enrich 3D visualization

**Yang Liu[1,†], Vinod Kumar Singh[1,†] and Deyou Zheng [1,2,3,*]**

[1]Department of Genetics, [2]Department of Neurology and [3]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Associate Editor: Alfonso Valencia

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Received and revised on March 2, 2020; editorial decision on May 9, 2020; accepted on May 12, 2020

## Abstract

**Summary:** Visualization in 3D space is a standard but critical process for examining the complex structure of high-dimensional data. Stereoscopic imaging technology can be adopted to enhance 3D representation of many complex data, especially those consisting of points and lines. We illustrate the simple steps that are involved and strongly recommend others to implement it in designing visualization software. To facilitate its application, we created a new software that can convert a regular 3D scatterplot or network figure to a pair of stereo images.

**Availability and implementation:** Stereo3D is freely available as an open source R package released under an MIT license at https://github.com/bioinfoDZ/Stereo3D. Others can integrate the codes and implement the method in academic software.

**Contact:** deyou.zheng@einsteinmed.org

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

'A picture is worth a thousand words;' all researchers know this and understand a well-prepared figure helps get messages across. Data scientists often use 3D figures to represent high-dimensional big data, e.g. those from thousands of samples or from simultaneous profiling of thousands or tens of thousands of molecules (genes, proteins and so on). Stereoscopic image, often used by structural biologists to show the 3D structure of a protein, DNA or large molecule complex, surprisingly has not been adopted by data scientists, despite its obvious advantage in illustrating the structure or relationship of the underlying data points. For example, the stereo view of a protein's 3D structure provides critical details on the back bone and side chain orientation of individual amino acid residues, facilitating better comprehension and interpretation of residue–residue interactions.

When plotted in 3D, the structure of many high-dimensional data actually becomes very similar to a protein molecule, as the points are like atoms while the connections can be considered as bonds between atoms. As a good example, let's look at single-cell analysis, in which investigators use single-cell RNA sequencing (scRNA-seq) to profile the transcriptomes of thousands to millions of cells (Stegle *et al.*, 2015) or single-cell assay for transposase-accessible chromatin sequencing (Satpathy *et al.*, 2019). After data acquisition and processing, dimension reduction is applied for the analysis and essentially for visualizing the otherwise unimageable cell–cell relationships. Principal component analysis is the conventional and most commonly used method for the task, but *t*-distributed stochastic neighborhood embedding (Amir *et al.*, 2013; Van Der Maaten and Hinton, 2008), uniform manifold approximation and projection (Becht *et al.*, 2019; McInnes *et al.*, 2018), Diffusion Map (Coifman *et al.*, 2005) and other algorithms have become more widely used to show cell clusters in 2D or 3D space. There, the clusters are determined by cells' similarities in their gene expression profiles. When the number of cells is small and the clusters are very distinct, regular 2D or 3D is sufficient. As the numbers of cells and clusters increased, even 3D representation may become overcrowding and looking at the data from different perspectives is necessary.

We demonstrate below that applying stereo imaging technology to the 3D plots can significantly overcome these limitations and deliver a much clear and intuitive message about cell–cell relationship.

In Figure 1A, we plot the clustering of the scRNA-seq data that we obtained from a human cerebral organoid (unpublished data), containing ~10 000 cells in 10 clusters. It shows that cells in some clusters overlap in the 2D projection, but additional views from a different perspective or modifying the projection may resolve the overlap. Plotting the same data in 3D stereo images significantly enhances our appreciation of the cell–cell relationship and heterogeneity (Fig. 1B). If the stereoscopic layer of visual separation seems less obvious for helping resolve cell heterogeneity in a single sample, its advantage will emerge more clearly when a fourth dimension is added, arising from comparative and integrative scRNA-seq analysis of multiple datasets. Visualization also augments data interpretation, such as in cell trajectory analysis, in which individual cells are projected to a pseudotime space representing the sequence of gene expression changes that a cell may go through in a dynamic biological process (Trapnell *et al.*, 2014), e.g. differentiation. As shown in Figure 1C and D, stereo view can significantly enhance our

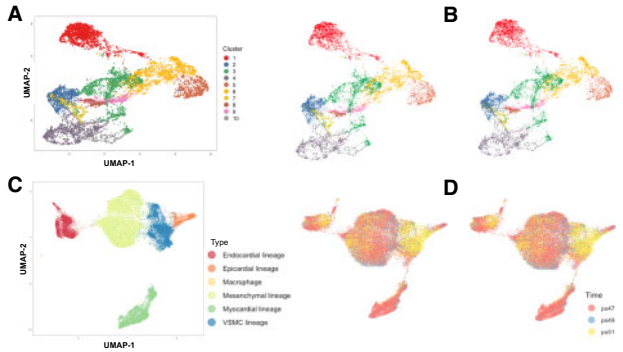


**Fig. 1.** Single-cell RNA-seq data in stereo view. (**A**) 2D UMAP plot of a brain organoid scRNA-seq data with 10 clusters. (**B**) Stereo view of (A), with the left- and right-eye images differing in about 5° in the vertical axis, indicating heterogeneity and subpopulations especially in the clusters 3 and 6. (**C**) 2D UMAP plot of single cells from the embryonic cardiac outflow tract, colored by cell types. (**D**) Stereo view of (C), colored by developmental stages, ps47 (47 pairs of somites), ps49 and ps51

appreciation of the dynamic cell trajectories, showing that both mesenchymal and cardiac lineages contribute to arterial vascular smooth muscle cells in the development of mouse embryonic cardiac outflow tract (Liu *et al.*, 2019). These examples (and additional ones in supplemental figure) clearly demonstrate the values of stereo imaging, but moreover one can further improve the visualization impacts using rotated or better animated images.

The preparation of stereo images in most cases is easy and certainly much easier than making videos. The pair of the left- and right-eye images are just two slightly different perspectives of the same image, e.g. by rotated along the vertical ($y$) axis in $\pm 3°$ (toward the readers) and also separated by the average interocular distance (~63 mm), as what were done to make the above figures. Because these steps can be incorporated to any 3D figure preparation software, we recommend data scientists to include them to take advantage of the power and effectiveness of stereo imaging for presenting big data. Nevertheless, to facilitate this, we provide an R package, Stereo3D, which is freely available at https://github.com/bioinfoDZ/Stereo3D. The software renders a user provided 3D data (e.g. UMAP) and a slightly different perspective of the same data into a pair of 3D images in order to achieve a 3D illusion of the object. Stereo3D can be used in a static mode to generate the plots or in an interactive mode with rotating and zooming (distance between pairs will not be set). In brief, the original set of data coordinates $(X,\ Y,\ Z)$ is rotated (in counter-clockwise direction) by an angle $\theta$ along $Y$-axis using a rotation matrix $R_y(\theta)$, so that the new set of coordinates is defined as

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = R_y(\theta) \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$= \begin{pmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

In summary, we believe that 3D stereo representation of high-dimensional data is a simple, elegant and effective way that allow researchers to better appreciate the data structure and help interpretation. Not only genomic data but also any 3D image can become more informative if presented in stereoscopic view, including those for protein interaction networks and neural or brain connectomes. We strongly encourage other investigators to include such images in their presentations and reports to harness the power of data visualization.

## Funding

This study was supported by National Institutes of Health grants [HD092944, HD070454, HL133120].

*Conflict of Interest*: none declared.

## References

Amir,e-A. *et al.* (2013) viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.*, **31**, 545–552.

Becht,E. *et al.* (2019) Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.*, **37**, 38–44.

Coifman,R.R. *et al.* (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. USA*, **102**, 7426–7431.

Liu,X. *et al.* (2019) Single-cell RNA-Seq of the developing cardiac outflow tract reveals convergent development of the vascular smooth muscle cells. *Cell Rep.*, **28**, 1346–1361.

McInnes,L. *et al.* (2018) UMAP: uniform manifold approximation and projection. *J. Open Source Softw.*, **3**, 861.

Satpathy,A.T. *et al.* (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.*, **37**, 925–936.

Stegle,O. *et al.* (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*, **16**, 133–145.

Trapnell,C. *et al.* (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, **32**, 381–386.

Van Der Maaten,L. and Hinton,G. (2008) Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.*, **9**, 2579–2605.