Visual Informatics 2 (2018) 14–25

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

MultiSciView: Multivariate Scientific X-ray Image Visual Exploration with Cross-Data Space Views

Wen Zhong[b,1], Wei Xu[a,*], Kevin G. Yager[a], Gregory S. Doerk[a], Jian Zhao[c], Yunke Tian[e], Sungsoo Ha[a], Cong Xie[b], Yuan Zhong[d], Klaus Mueller[b], Kerstin Kleese Van Dam[a]

[a] *Brookhaven National Laboratory, Upton, New York, United States*
[b] *Stony Brook University, Stony Brook, New York, United States*
[c] *FX Palo Alto Laboratory, Palo Alto, California, United States*
[d] *Facebook Inc., Menlo Park, California, United States*
[e] *Midea Emerging Technology Center, San Jose, California, United States*

ARTICLE INFO

Article history:
Received 11 December 2017
Received in final form 23 February 2018
Accepted 12 March 2018

Keywords: Scientific x-ray image visualization, Multi-level scatterplot, Cross–data space exploration

ABSTRACT

X-ray images obtained from synchrotron beamlines are large-scale, high-resolution and high-dynamic-range grayscale data encoding multiple complex properties of the measured materials. They are typically associated with a variety of metadata which increases their inherent complexity. There is a wealth of information embedded in these data but so far scientists lack modern exploration tools to unlock these hidden treasures. To bridge this gap, we propose *MultiSciView*, a multivariate scientific x-ray image visualization and exploration system for beamline-generated x-ray scattering data. Our system is composed of three complementary and coordinated interactive visualizations to enable a coordinated exploration across the images and their associated attribute and feature spaces. The first visualization features a multi-level scatterplot visualization dedicated for image exploration in attribute, image, and pixel scales. The second visualization is a histogram-based attribute cross filter by which users can extract desired subset patterns from data. The third one is an attribute projection visualization designed for capturing global attribute correlations. We demonstrate our framework by ways of a case study involving a real-world material scattering dataset. We show that our system can efficiently explore large-scale x-ray images, accurately identify preferred image patterns, anomalous images and erroneous experimental settings, and effectively advance the comprehension of material nanostructure properties.

1. Introduction

X-rays are used modernly to probe the characteristics of materials. These experiments are frequently conducted at 'synchrotrons'—circular electron accelerators used to generate bright and collimated x-ray beams. The different experimental types and associated measurement modalities yield different kinds of images, which provide various information about the structure of a probed material. These x-ray images are distinct from natural (visible-light) images since they do not directly describe the physical appearances, but instead capture far-field attenuation, diffraction, or scattering patterns of the materials. Such images can appear quite similar upon casual inspection, even for materials whose structures are very different. Subtle differences in these images can be hard to observe since they are high-dynamic-range grayscale data whose features are only recognizable to trained experts. Therefore, conventional image browsing tools—typically targeted towards photographs—provide only a very rudimentary view of the dataset.

Another challenge in working with these images is rooted

*Corresponding author: E-mail address: xuw@bnl.gov
Homepage: https://www.bnl.gov/compsci/people/staff.php?q=113
[1] This work was done when Wen Zhong was a Ph.D. student supervised by Dr. Wei Xu and Prof. Klaus Mueller at Stony Brook University and Brookhaven National Lab.

https://doi.org/10.1016/j.visinf.2018.04.003

in their heterogeneous nature. For instance, each x-ray image can also be associated with multiple attributes, including its chemical or physical properties, analysis parameters, and experimental metadata. In scientific experiments, these attributes usually depend on each other and thus are *multivariate*. The acquired data can be a mixture of nominal, numerical, categorical, and spatiotemporal data. Beyond that, modern scientific experiments are generating data at an increasingly rapid pace. Altogether, it challenges the visualization and exploration of large scientific datasets by the domain expert to achieve complete understanding of the full dataset scope, and apply the analysis appropriately. Currently, there is no visualization work focusing on this type of analysis and data challenge. The practitioners can only visualize these data separately by collecting a set of meaningful plots manually and applying association tediously. The deficiency of modern exploration tool greatly hinders the understanding of scientific x-ray data toward new findings.

To bridge this gap, in this work, we propose *MultiSciView*, a multivariate scientific x-ray image visualization system to remedy this missing capability. In general, a set of visualization and exploration methods is devised to allow scientists to easily interact with both raw data (hereafter *image space*) and analysis results (hereafter *attribute space*) in a single context, and thereby validate and understand important conclusions. To be specific, we propose a two-step methodology for the exploration of experiment image datasets. Firstly, multi-level image exploration and attribute relation projection are enabled for efficient large-scale image space and attribute space understanding respectively. Then, the image and attribute spaces are unified using collaborative filtering for coordinated exploration of experiment dataset.

In order to evaluate our method, we adopt x-ray measurement scenario: *x-ray scattering*, a popular and powerful technique for measuring the physical structure of materials at the molecular and nano-scale, as case studies. However, our methodology is not limited in the scope of x-ray scattering. In fact, for most multivariate scientific x-ray images—x-ray medical images, x-ray diffraction images and electron energy-loss images [3]—where image exploration plays an essential role for understanding, our method is also applicable. As far as we know, there is no other visualization work optimized towards exploring and understanding large x-ray scattering datasets. On the whole, our paper offers the following contributions:

- We propose a multi-level scatterplot, providing the visualization of images in overview, image and pixel levels as three different scales through zooming.
- We utilize a Pearson correlation based attribute similarity visualization for high level material comparison and understanding.
- Using cross filters, we unify image and attribute spaces, and obtain the opportunity of coordinated exploration across data spaces.
- Through performing real world case study with domain scientists, we demonstrate that our system could efficiently explore the experiments, find erroneous and anomalous data, and advance material understanding.

The remainder of the paper is structured as follows: Section 2 reviews the background and related works. The scattering image dataset and the general framework of our method are described in Section 3 and Section 4 respectively. Section 5 and 6 propose the major visualizations of image and attribute spaces for scattering image understanding. A real-world case study is examined in Section 7. Finally, the paper is concluded in Section 8 with a brief discussion of limitations and future directions.

2. Background

In the past decade, visualization has made profound advancements, especially for high-dimensional data, where a variety of works have been introduced covering all three stages of the visualization pipeline—from data transformation, to visual mapping and view transformation, to integrating interactivity [19]. Similarly, for scientific data, visualization also has demonstrated its effectiveness to explore, analyze and gain insights into such data [11]. The state-of-the-art visualization methods dealing with single facets of scientific data, are well-established [16].

When the scientific experiments get more complicated, the categories of scientific data are more completely summarized to "multifaceted" data and model scenarios for modern experiments [16]. There are five major categories: 1) spatiotemporal data, 2) multivariate data consisting of different attributes, 3) multimodal data stemming from different sources, 4) multirun data from multiple simulation runs, and 5) multimodal data from multiphysics simulations of interacting phenomena. However, in spite of the existing works, increasing experimental and data complexities still challenge current tools, necessitating the development of new advanced approaches. In the following subsections, we will review related works and also discuss the background of the x-ray scattering scenario.

2.1. Multi-attribute data exploration

There are a number of methods supporting multi-attribute data exploration, such as exploring large faceted collections. Early systems FOCUS [27] and InfoZoom [26] support dynamic queries in the format of a large table with pivoting, expanding and other operations on the attributes. Later, [31] introduced an image searching tool based on progressive queries by filtering metadata. However, these techniques have fixed table-wise layout and rely on linear exploration process that requires users step by step refinement with increasing specifications.

Another group is based on scatterplot layout—the most commonly used visual tool for multi-attribute data, since it is straightforward to understand, and familiarized by the majority of communities. It has been used as a standard 2D display approach [17] where users can easily alter the attribute combination. An improved approach called Flow-based Scatterplots aims to augment scatterplots by using sensitivity coefficients to

highlight local variation of one variable with respect to another [5]. Therefore, the flow variations resemble the variation correlations between two variables. Nevertheless, the visualization is limited to only pairwise comparison. Scatterplot matrix [6] is the extension of scatterplot where a full map of bivariate scatterplots are organized in a matrix. Consequently, the pairwise attribute correlation can be observed simultaneously among all the attributes. But when the dimension increases, it becomes more difficult to explore. In order to provide a clear structure to the visual exploration process, Rolling the Dice was introduced as an interactive navigation in the scatterplot matrix space [10]. The transitions among selected scatterplots can be seamlessly integrated; thus both comparison and correlation can be conveniently conducted. However, the major downside of these approaches is that the visualization resides only in the attribute space and does not suite to explore cross-data space exploration that fuses other data spaces into a single context.

2.2. Coordinated multiple views

As reviewed in [24], coordinated multiple views have been steadily developing and shown to be effective in exploratory visualization. Different perspectives of the dataset are visualized, explored and analyzed in multiple linked views to compare simultaneously. These views usually support interactive feature specification via brushing for multivariate data exploration.

XmdvTool [29] is comprised of a set of baseline visualization tools such as scatterplots, parallel coordinates, hierarchical techniques, and so on, to visualize multivariate data with linked N-dimensional brushing. Similarly, GGobi [28] supports linking of views and various types of visualization and interaction, with additional methods extended from Grand Tour [1]. For scientific data as summarized in [16], SimVis [9], WEAVE [14] and PointCloudXplore [25] are three representative visual analysis systems. They link attribute views with physical views of volumetric data so that the attribute can be connected with spatial context. Similarly, [2] presented a linked physical and feature space views for multi-field medical data. However, these works lack of multiple levels of details during exploration, which is an important characteristic in examining large-scale heterogeneous datasets.

2.3. Visual comparison at the image level

Visual comparison of complex objects is an important component involved in data analysis. A survey of work in information visualization related to comparison was presented by Gleicher et al [12] [13]. As indicated in their papers, although there is a great diversity of systems and approaches, all designs are assembled from the building blocks of juxtaposition, superposition, and explicit encodings. Each of them has its superiorities and tradeoffs. Bearing in mind these basic forms of visual design with both factors would lead to transfer designs to applications, and develop novel visual depictions.

For the side-by-side image comparison, a number of works for scientific analysis fall into the aforementioned building blocks. For instance, [15] arranges the visual comparison of volume rendering results from different view positions or transfer function settings in a spreadsheet layout. But this fixed layout is limited to connect to other attributes. In contrast, a multi-image view visualization [21] is designed to support the comparison of series of scans from the same specimen simultaneously in a hexagon tile pattern. This visualization is useful to reveal detail differences of the scans while still keeping the sample space unchanged. However, the pieces of scan images are stitched together to form a complete picture, thus a full image comparison is not possible.

2.4. X-ray scattering image

In the x-ray scattering experiment, a narrow and collimated beam of x-rays is directed through a sample of interest, and one records the far-field pattern of scattered x-rays on a two-dimensional detector. This far-field pattern arises from the interference of the innumerable microscopic scattering events occurring within the material. The detector image thus encodes the materials physical structure; e.g. with bright spots appearing when atoms/molecules form regular arrays (diffraction from a crystal lattice).

Like other x-ray images, each scattering images from beamline experiments have a large amount of associated metadata, which we can classify into three main types: (1) a definition of the material being studied, both in terms of its composition (chemical makeup, fabrication history), and the conditions under which it was measured (temperature, pressure, etc.); (2) the experimental parameters used for the measurement (beam energy, exposure time, etc.); (3) results of data analysis on the image (statistical metrics, peak fitting, etc.), which can be converted into physical insights about the material (molecular packing motif, crystal size, etc.).

Due to the data complexity, a key challenge is to provide the experimenter with the flexibility to control the data layout. Different materials and experiments will involve remarkably different sets of both materials parameters and data analysis metrics; moreover, oftentimes the most relevant parameters are not known before the experiment is conducted, and must instead be identified by exploring the dataset. Experimenters must have both high-level summaries of the trends from data analysis, as well as easy access to the corresponding raw data, in order to confirm the results and tweak data analysis as necessary. However, to the best of our knowledge, there is no visualization method optimized for the exploration of this kind of datasets.

3. X-ray Scattering Image Dataset

In this work, we focus on visualization of x-ray scattering experimental data, which is a series of images collected with numerical attributes and metadata [22]. There are 79 different samples, categorized as *C67* and *L74* [23], two kinds of polymer materials. These samples were organized into well-defined nanostructures [4, 7], exposed to heat (‘annealing’) [20] and held at a specific temperature. Then they were scanned by x-ray beam and a time sequence of scattering images were collected in the far field. They are the raw data for the analysis of material nanostructures. The presented experiments focused on studying the annealing nanostructure order; that is, the gradual improvement of the order in the samples as they are heated [20].

In particular, the effects of annealing temperature and time, material thickness, and the blended polymers (two kinds of polymers: *S3M3* and *S6M6*) were investigated with respect to how they influence the ordering behavior [30] and the ordering rate [8] of the materials.

Fig. 1(a) shows one example x-ray scattering image, where two bright spots ('scattering peaks') appearing on either side in yellow rectangles arise due to the well-ordered nanostructure. For the other peaks or structures, such as horizontal or vertical bands, they are caused by the occlusion of beamline equipment and hence not considered. By contrast, the blurred scattering image in Fig. 1(b) displays two blunt spots because of bad nanostructure. Associated with one image, there is a set of quantitatively analyzed (not all the images are analyzed) attributes, shown in Fig. 1(c). Among these attributes, the peak width σ_1, as a key metric, is highly related to the quality of the image. In general, a sharp peak (small σ_1), as a well-ordered nanostructure, is preferred. Whereas a broad peak (large σ_1) shows poor nanostructure orders. Therefore, highly-ordered materials (which exhibit clear, sharp and bright peaks Fig. 1(a)) are desirable and vice versa (Fig. 1(b)).

4. System Overview

In this work, we closely collaborate with the domain scientists, and design our system based on their requirements and suggestions. The two material scientists (also co-authors) are professional material scientists especially for nano-structures and structural characterization areas for more than 10 years. They mainly adopt beamlines experiments to refine or design new materials with better nanostructures.

By considering the domain scientists exploration preferences, the aforementioned characteristics of the experiments and the challenges of understanding the scattering images, we summarize users' major requirements as follows:

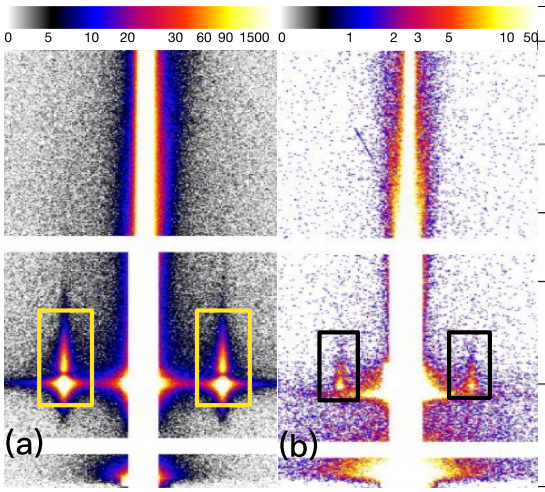
- **(R1):** Compare material comparison through effective arrangements of all the scattering images.
- **(R2):** Understand high level understanding of various attribute relations for different materials.
- **(R3):** Identify anomalies, including operation errors and experimental settings, for better data quality.
- **(R4):** Explore images in different scales, including checking detail pixel intensity.

Therefore, we devise a system, *MultiSciView*, shown in Fig. 2, to efficiently visualize the scattering images by integrating image and attribute spaces together in a multiple level-of-detail mechanism, as well as linking with other coordinated views. There are three major visualizations: 2.b *multi-level scatterplot*, 2.a *cross filters* and 2.c *attribute projection*. Specifically, the main visualization *multi-level scatterplot* (b) is the image space visualization adopting scatterplot layout (**R1**) with three levels of details (**R4**), which enables the user to obtain general understanding of the whole experimental space. Since each point in the scatterplot represents a single image, when visualizing images instead of scatterplot points (**R2**), the pairwise attribute comparison is integrated with the image space. Then a side-by-side visualization can be realized naturally (**R3**). As auxiliary visualizations, the exploration of attribute space consists of a histogram based *cross filter* (2.a) visualization and an *attribute projection* (2.c) visualization. Linked to the multi-level scatterplot, these two components play as collaborative visualizations, revealing potential relations among attributes with regards to different materials (**R2**) and identifying anomalies (**R3**). Based on the above overall description of the system, we plan to elaborate our image and attribute space visualizations in the following two sections.

5. Image Space Visualization

The image space visualization is comprised of a multi-level scatterplot and an auxiliary image detail visualization as shown in Fig. 2(b). It is indeed an integrated visualization of image space and the attribute space. During our discussion with the domain scientists, we proposed many other visualization prototypes, such as stream graph and glyph-based line charts. We found out a more simplified and user familiar layout is more desirable to enable the exploration of large-scale scattering images by domain scientists. Therefore, the design with a traditional scatterplot layout with additional three levels of details is more convincing. In specific, the major visualization is a standard scatterplot where the user could select any two attributes for pairwise attribute comparison. In addition, we encode a third attribute into this layout by color coding each scatterplot point. As shown in Fig. 3(a), x and y axes are *annealing_time*, *sample_name*, and z axis encodes σ_1 using *inferno* colormap. We tried many other color schemes. However, *inferno* colormap color range is the most clear and distinctive and is preferred by users (Fig. 3). This layout provides the overview level of details. In this example, the scientist can check the experiments listed along sample names in the y axis. Due to the large amount of scattering images and the limited analysis ability, not all of the images are analyzed before exploration. The unanalyzed samples thus have only metadata attributes. To label these data, since we allow users to choose any color that is exclusive from *inferno* color scheme, for these samples, green color is chosen by users to distinguish from analyzed images.

When the scientist would like to zoom in for the details of the scattering images distribution, Fig. 3(b) would be shown firstly to visualize image evolution with respect to the changes of experimental settings for detail comparison. In this case, each row represents a series of images for a sample. Continuing to delve into details, Fig. 3(c) would be displayed as the second level, where each point evolves into the corresponding scattering image. To enable a close look of the image detail without losing the context, we attach an auxiliary visualization to show the single image (Fig. 3(d)). This visualization can be further zoomed in to check more details until it reaches the final level as in Fig. 3(e), where interested pixel values are superimposed for quantitative analysis.



Attribute	Description
Annealing temperature	The temperature the sample was heated to.
Annealing time	The amount of time the sample was held at high-temperature.
fit_peaks_prefactor (prefactor)	The intensity of the measured scattering peak. (Higher intensity indicates more ordered material.)
fit_peaks_d_0 (d_0)	The size-scale of nanostructures within the material (computed based on the scattering peak position).
fit_peaks_σ_1 (σ_1)	The width of the measured scattering peak. A large width (broad peak) indicates poor order, while a small width (sharp peak) indicates good order.
fit_peaks_m (m) and fit_peaks_b (b)	Fit for the amount of background scattering (linear model), which is a measured of the amount of disorder (e.g. roughness) in the sample.

(c)

Fig. 1. (a) An example of 'good' x-ray scattering image. The two sharp and bright peaks in yellow rectangles, arise due to distinct nanostructures, which is preferred. (b) An example of 'bad' x-ray scattering image with blurred background and blunt peaks, which is not preferred. The color legends are in the front of these two images. (c) Table enumerating some attributes relevant to the experiments discussed in this work.

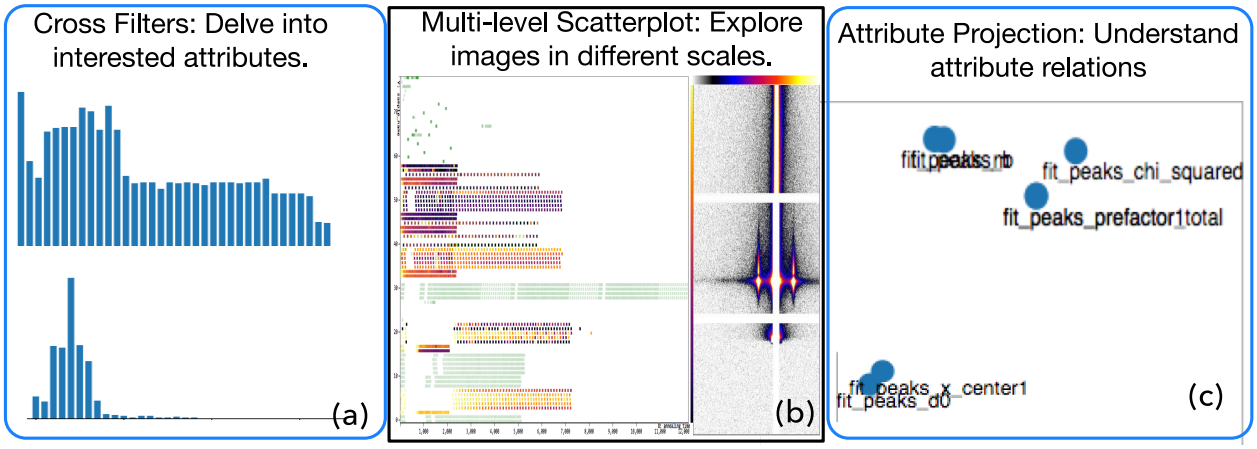


Fig. 2. *MultiView* overview includes (b) image space visualization by a scatterplot for multi-level image exploration; (a) and (c) attribute space visualization for collaborative exploration and understanding.

6. Attribute Space Visualization

To understand attribute relations, the attribute space visualization is composed of two visualizations: attribute correlation projection and histogram based cross filter, where the former focuses on the general understanding of relations among attributes, while the latter allows the interactions between the image space and the attribute space.

6.1. Attribute Projection

Since the relations of attributes reveal materials physical properties, the user would like to explore the correlations of these attributes in order to better understand materials properties. Beyond the scatterplot exploration in the image space visualization that reveals pairwise correlations, a global relationship among attributes is important as well. Therefore, we construct an attribute representation based on the attribute values of scattering image, and project them into a 2D space to show the potential relations. For the distance metric, after trying and comparing many popular distances such as Euclidean distance and Cosine distance, Pearson correlation is employed as user preferred metric. For these attributes, the expert would like to fully understand their relations to verify experiments. Pearson correlation captures pair-wise linear positive or negative relations and helps the user understand the attribute relation pair by pair. Specifically, we provide the metric as: $1 - |Pearson(x, y)|$. Since the correlation range is between -1 and +1, this metric would capture both positive and negative correlated relations. Inspiring by [32], we employ a color coded multidimensional scaling (MDS) 2D embedding method to display attribute relations. As shown in Fig. 4, 4.A and 4.B visualized the attribute relations of both materials *C67* and *L74* respectively. To be specific, green and red lines are used to denote positive and negative relations respectively. Besides, we adopt stroke opacity to encode the correlation weight. In this way, the user could easily obtain correlation information. We will demonstrate its use in the case study section.

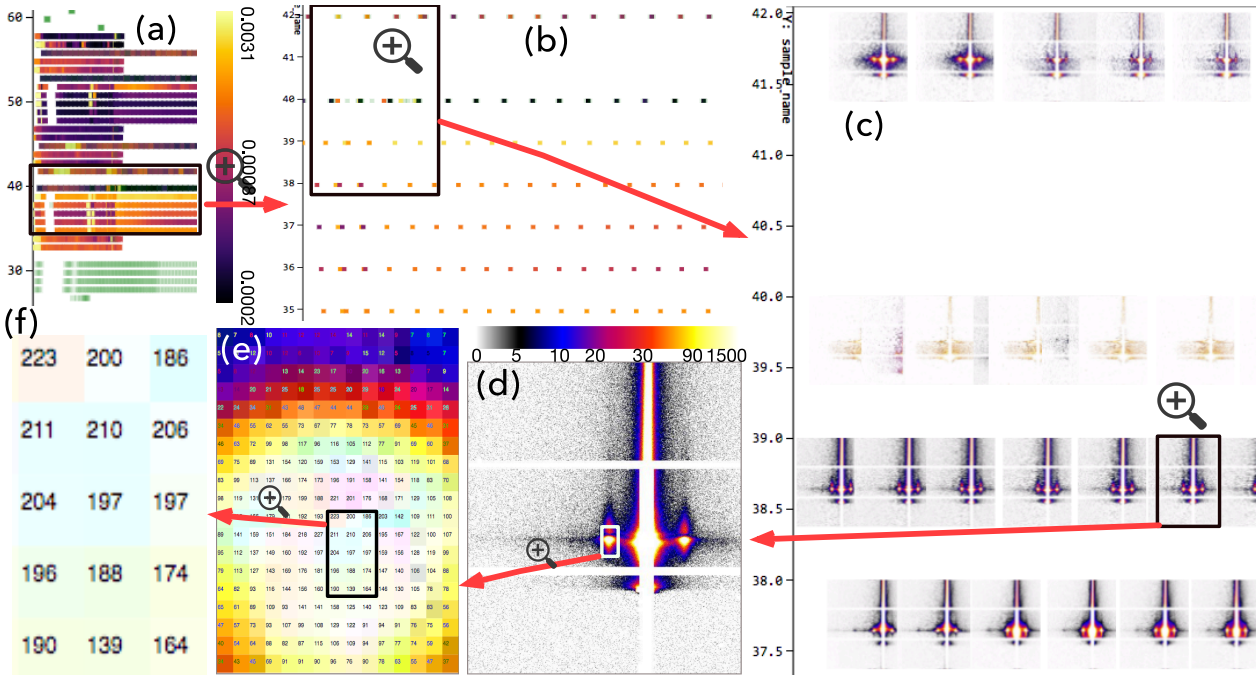


Fig. 3. The pipeline of our image view, a multi-level scatterplot. Through zooming into different levels, it supports exploring scattering images in three different levels of details. (a) First level, overview arranged by attribute orders. (b) Zooming in first level to check details. (c) Second level, delving into image groups. (d) Zooming in second level to observe one image. (e) Third level, reading specific pixel values. (f) Exact value numbers.

6.2. Attribute Cross Filters

In many cases, the user exploration starts with a single attribute because of more familiarity, so the interaction between attribute and image spaces is an effective way to find further relations among different samples. In order to enhance such exploration, a group of histogram based cross filters are provided. The interface not only provides the histogram of each attribute, but also allows the user to apply the knowledge achieved along any filter to merely visualize the attributes and the corresponding scattering images of interest in the multi-level scatterplot visualization. The histogram could be visualized as bar chart or stacked bar chart based on user's preferences. For instance, as shown in Fig. 5(A), there are four salient peaks in the histogram of values for d_0; the user can brush to interact with this data and select a particular peak in the distribution, consulting the image visualization to understand the origin of this part of the distribution. We will also evaluate its use in the case study part.

7. Case Study

We perform a case study with two material scientists, who derive the design requirements, to evaluate our methodology. Their feedback is used to show whether our system could effectively help them identify the potential properties of materials, and steer subsequent experiments. Based on the preferred image scattering patterns and important metrics introduced in Section 3, we describe our case study in the following subsections.

7.1. Material General Understanding

Since multi-level scatterplot is the key part of our system, we first present several use cases for dataset general comprehension using this visualization. Before introducing the details, we first list some important visualization patterns that are useful for users' justifications: 1). Since smaller σ_1 usually indicates better nanostructure, so darker colors in scatter plot are preferred. 2). Material nanostructure is usually stable, so with the increase of *annealing* temperature, the changes should be continuous. In this way, the colors in scatter plot should change smoothly without color jumping. 3). As mentioned in Sec. 3, patterns that are diagnostic of "good" samples should be focused. The following tasks will be finished based on these patterns.

7.1.1. Material Exploration and Comparison

For a general understanding of the scattering images, the user first adopted the multi-level scatterplot to explore the whole set of images. Since the principal aim of the experiment was to monitor scattering images ordering over time, the user selected to visualize with *annealing_time* along the x-axis, and *sample_name* along the y-axis, with the z-axis (color of scatterplot points) used to denote the value of σ_1. As shown in Fig. 6(A), an *inferno* color scheme is used to denote the z-value, and a green color denotes unanalyzed samples. As shown in the blue rectangles, there are three main categories of materials: 6.a: *L74*, 6.e *C67* and 6.i *MB*, where the former two are analyzed in detail.

Considering σ_1 as a useful metric, the user targeted on *L74* samples 6.b (names 18 – 22) and *C67* samples 6.f (names 42 – 58) images with small σ_1 values (darker colors). However, he noticed that the z-axis colors of 6.f samples did not

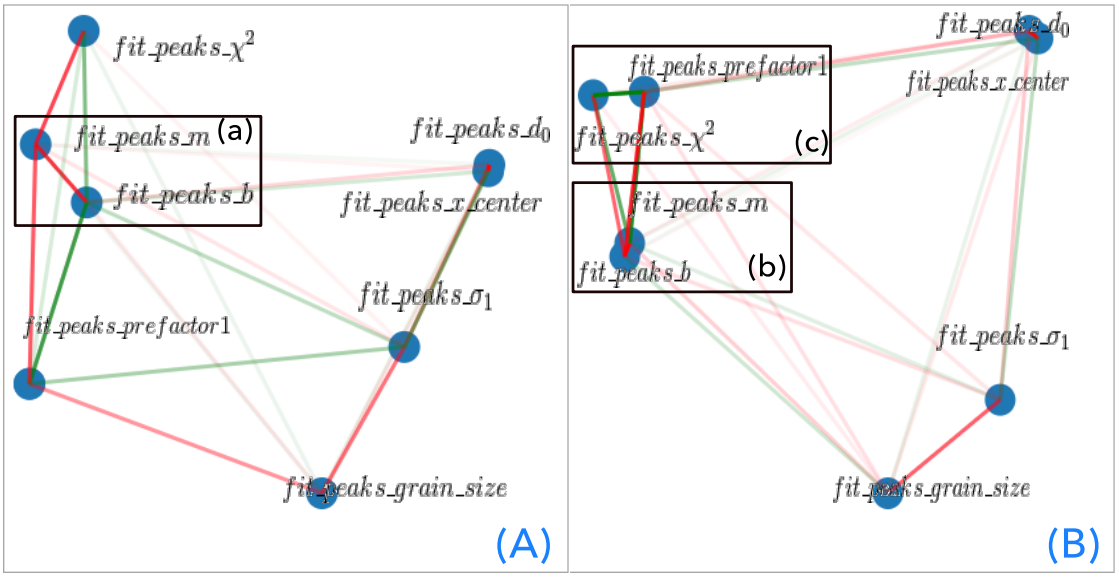


Fig. 4. Attribute correlation projections: (A) *C67* material attribute projection, (B) *L74* material attribute projection. The attributes in rectangles reveal differences between *C67* and *L74*.

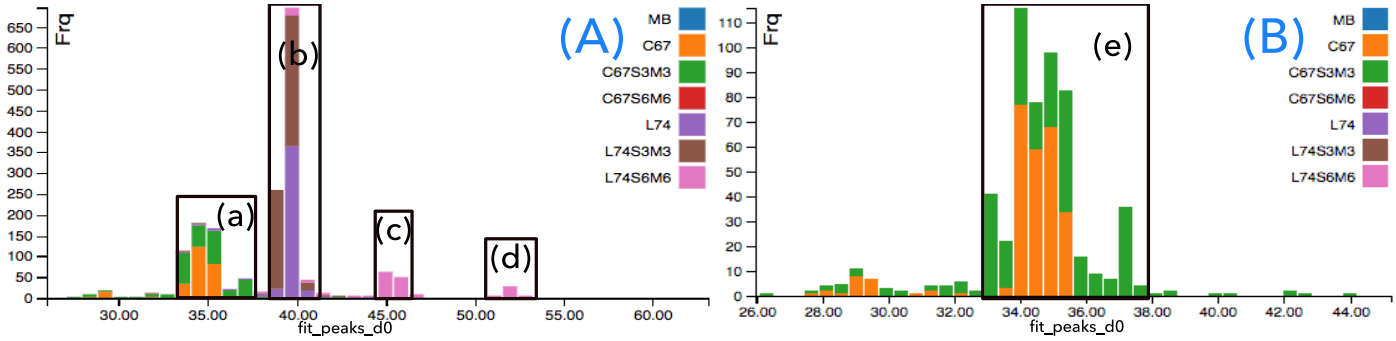


Fig. 5. *Fit_peaks_*d_0 attribute stacked histogram based cross filters. (A) Shows both *C67* and *L74* materials. (B) Only shows *C67* material.

exhibit a smooth trend, meaning that the σ_1 values fluctuated substantially over time. This could be the indicative of instable nanostructures, or poor data-fitting (leading to spurious analysis results). To further check the samples detailed nanostructures, he zoomed into the corresponding next level raw images 6.k and found the images were blurred (weak low signal-to-noise) and exhibited spurious features, such as blunt peaks or single streak near the origin, instead of a clear scattering peak. Thus, the user was able to rapidly identify that this sequence of experiments was unsuccessful (likely due to a misalignment of those particular samples during measurement), and that this set of analysis results should not be pursued. This set of experiments was thus scheduled for being rerun.

Subsequently, the user investigated 6.b samples, which exhibit a smooth changing trend of σ_1. The user immediately zoomed-in to the raw image series 6.j to check whether these samples were well-aligned during measurement. Except for several anomalous images showing blurred backgrounds, most of them displayed well-ordered nanostructure. He also toggled to pixel display mode for the examination of detailed pixel values and confirmed the peak values were very large. These samples were thus identified as valuable for further study.

7.1.2. *Blending Polymer Influence Investigation*

Using the zoomed-out scatterplot view, the user was then able to investigate one of the fundamental hypotheses of this scientific experiment: whether blending polymers into materials influences the nanostructure orders. The user first compared *L74* pure sample (names 40) with *S3M3* and *S6M6* blended samples 6.c, immediately recognizing that blended lead to a dramatic and consistent decrease in σ_1 (continuous darker colors). This indicates that blending greatly improves nanostructure order quality. The trend was confirmed by selecting representative samples, and zooming-in in order to compare their raw scattering images side-by-side, where the difference in peak width was immediately confirmed in the raw data. The peak intensities were also compared by zooming into the pixel-level for these representative samples, where it was found that the highly-ordered samples (*L74* with blending) had high peak intensities (with pixel values as large as several thousand). Overall, this exploration exercise allows the user to rapidly test and confirm one of the important scientific hypotheses.

After this discovery, the user returned to the *C67* material. The user browsed the remain analyzed data 6.h, finding the quality to be marginal, explaining the color fluctuations ob-

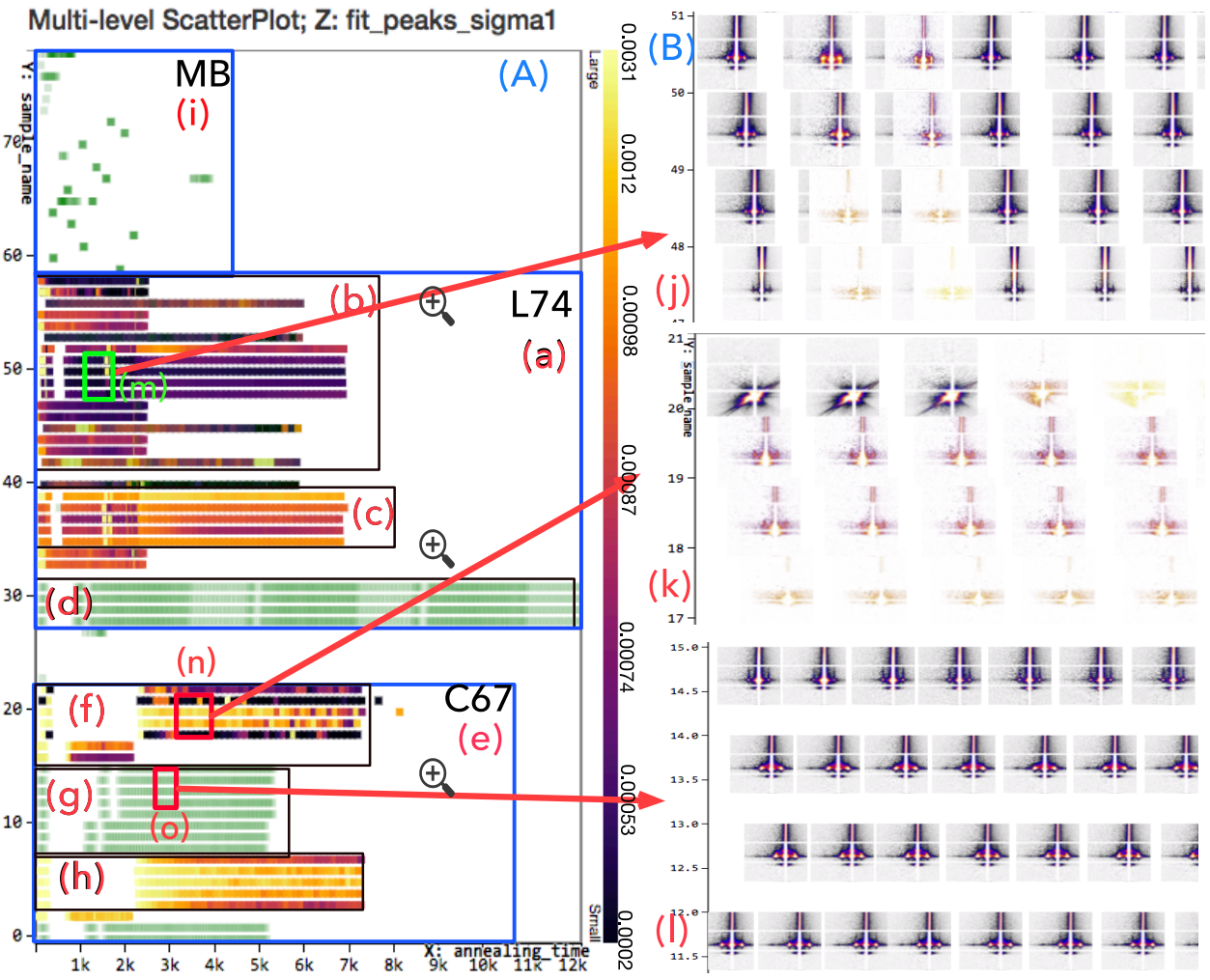


Fig. 6. Multi-level scatterplot. (A) The scatterplot of *annealing_time*, *sample_name* and *fit_peaks_*σ_1 attributes as x, y and z (meaning the point color) axes, respectively. A domain expert can immediately recognize that samples 35-39 (c) are 'good' in the sense that they exhibit a smooth trend (σ_1 decreasing with time), while samples 48-52 (part of (b)) are 'even better' in the sense that the achieve better order (smaller ultimate value of σ_1). (B) The corresponding image series when zooming in.

served in the trend. The user then investigated a set of unanalyzed samples 6.g, finding this data 6.l to be clear and highly robust. This observation agreed with previous conclusion that *C67* experiments were unsuccessful and thus were identified as high-priority for the next round of data analysis.

7.1.3. Anomalous Image Detection

During the exploration process, the user also identified some anomalous experiment results. For example, as shown in Fig. 7, the patterns for first four 7.b samples were abnormal. The color jumped from light yellow to blue two times, which was physically impossible, since physical properties of materials tend to change gradually in time. The user zoomed-in to investigate the raw images 7.c, and discovered that the first and third images did not have strong peaks, while the second and fourth had strong and sharp peaks. This was consistent with the colors of scatterplot points 7.b. Suspecting an experimental error, the user changed the x-axis to the *sequence_ID* attribute (shown in 7.d), which is opposite in that it increments with every experiment performed, and thus provides an unambiguous ordering of the experimental data collection 7.f. In this view, the trend of data was smooth and sensible. By comparing the ordering of the images in the *annealing_time* and *sequence_ID* views, the user was able to identify that the metadata labels of *annealing_time* were erroneous for a small set of images (owing to human error while collecting the data). Having identified the error, the labels could be fixed, and the dataset can be used for further analysis. Similarly, the user was able to identify gaps in the data collection 7.e, which are prime candidates for follow-up experiments in order to improve the overall data quality and the robustness of the measured nanostructure order trends.

7.1.4. Exploration of Experimental Plans

Besides the *annealing_time* ↔ *sample_name* scatterplot, the user also explored other scatterplots, such as *annealing_temperature* ↔ *annealing_time*. As shown in Fig. 8, visualization 8.A immediately summarized the experimental plans for the entire experimental run. One can identify a few experimental design patterns. Some samples were measured at room temperature (cluster of green data-points in (a)). However, most

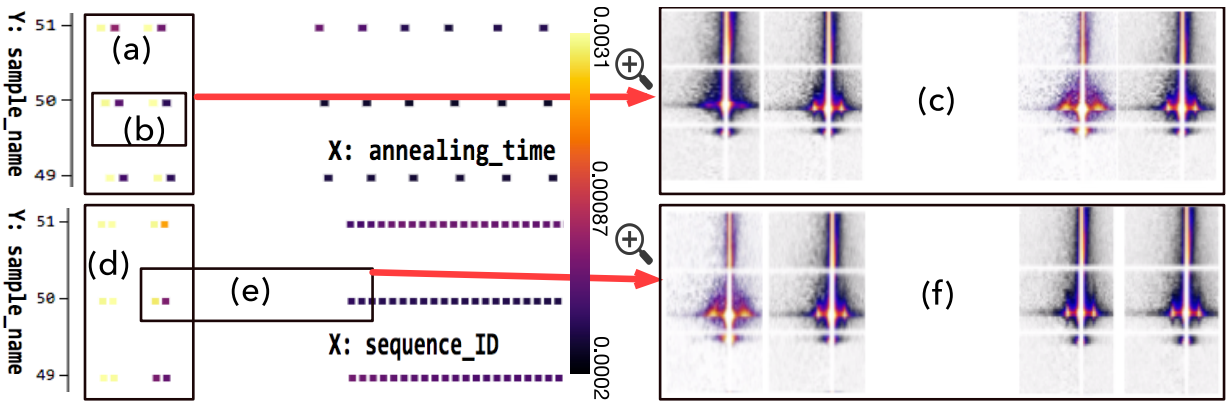


Fig. 7. For sample 51, the recorded *annealing_time* is erroneous. By comparing with the *sequence_ID* attribute, one can identify that an experimenter error led to mis-labeling of some of the images.

experiments involved increasing annealing temperature. Many annealing runs were performed by sharply increasing the temperature to 200 °C 8.b or 220 °C 8.c. In a more complex experiment, the user increased temperature to 200 °C 8.b, then 220 °C 8.d, and finally 240 °C 8.e. This could help check whether well-ordered materials could be further ordered at higher temperature. By zooming into raw images 8.f, 8.g, 8.h and 8.i, the user could confirm that increased temperature led to improved ordering in the materials, however, after 220 °C, the quality became worse. The user found this data-view to be a useful summary of the experimental execution, allowing one to recall what kinds of experiments had been performed.

7.2. Attribute Projection Analysis

After exploration of the multi-level scatterplot, the user exploited the attribute relation projection to look for high-level trends in the data. As shown in Fig. 4, 4.A and 4.B are the attribute correlation projection of *C67* and *L74* materials respectively. The user found that the differences between these two projections mainly lied in the correlations of m and b attribute pair, and *prefactor1* and χ^2 attribute pair. After observing this unexpected trend, he firstly switched back to the scatterplot in order to investigate the strong negative relation (red line) between m and b attributes in Fig. 9. It is obvious that the scatterplot of *C67* contains a large number of anomalous images 9.c with extremely large σ_1 values, whereas 9.b fits better to linear relations. The data quality for *C67* images was found to be low. In other words, because many of the *C67* experiments were unsuccessful, with poor sample alignment leading to noisy images, the correlation between m and b was weaker than it would be for high-quality data (such as for *L74)*. This confirmed the previously-noted conclusion (that some *C67* experiments would need to be repeated), while also established that m and b are strongly correlated.

By investigating further correlations in this manner, the user was able to uncover several unexpected trends. For instance, the background scattering (quantified through m and b) was found to be positively correlated with the intensity of the scattering peak (*prefactor1*). Conversely, d_0 and σ_1 are not strongly correlated, suggesting that the quality of order is not strongly dependent on the precise size-scale of the nanoscale ordering. Each of these trends can be investigated in turn using the scatterplot.

7.3. Attribute Collaborative Filtering

After exploring the attribute relations, the user would like to deeply delve into specific attribute. The d_0 attribute (peak position) encodes information about the size-scale of the nanostructure. As shown in Fig. 5, there are four peaks in 5.A histogram of d_0 values. The user surprisingly noticed that 5.a composes of *C67* and *C67S3M3*; 5.b composes of *L74* and *L74S3M3*; 5.c and 5.d compose of *L74S6M6*. *S3M3* and *S6M6* denote additional small and large molecule polymer materials blended into these samples respectively. He then zoomed into 5.a to check the detail distributions of *C67* material (5.e), where *C67* (orange color) and *C67S3M3* (green color) were still blended together, forming one pattern. Since the experimental conditions for studying *C67* and *L74* were very similar, the user was curious about why *L74* exhibits several well-defined d_0 value groups, while *C67* only forms one value group. To answer this question, after brushing these patterns, he observed 5.c corresponds to *L74S6M6hp50*, and 5.d corresponds to *L74S6M6hp70* (hpXX denotes the volume percentage of the blending). For 5.b, with the decrease of *S3M3* blending percentage, d_0 shifted rightward, mixing with pure *L74*. This is highly due to the fact that *S3M3* molecule mass is smaller than *L74*. From this exploration of the histogram, the user deduced some surprising conclusions: the *C67* material does not swell (change size) when blending occurs, whereas the *L74* material swells under certain conditions (especially when the amount of blending is large). This unexpected difference in materials response was interesting to the user, and was selected as an avenue for future experimental investigation.

In addition to exploring the trends in d_0, the user also explored the samples with largest *grain_size* attribute (where a large grain size indicates good order). As shown in Fig. 10, they first brushed 10.a samples and found that many sample patterns were similar to 10.c. These few samples with very large grain size were suspicious since they did not exhibit a smooth trend in the scatterplot color view. Further zooming revealed that the raw images were very noisy 10.e with no scattering peaks. These spurious results could thus be ignored. They then

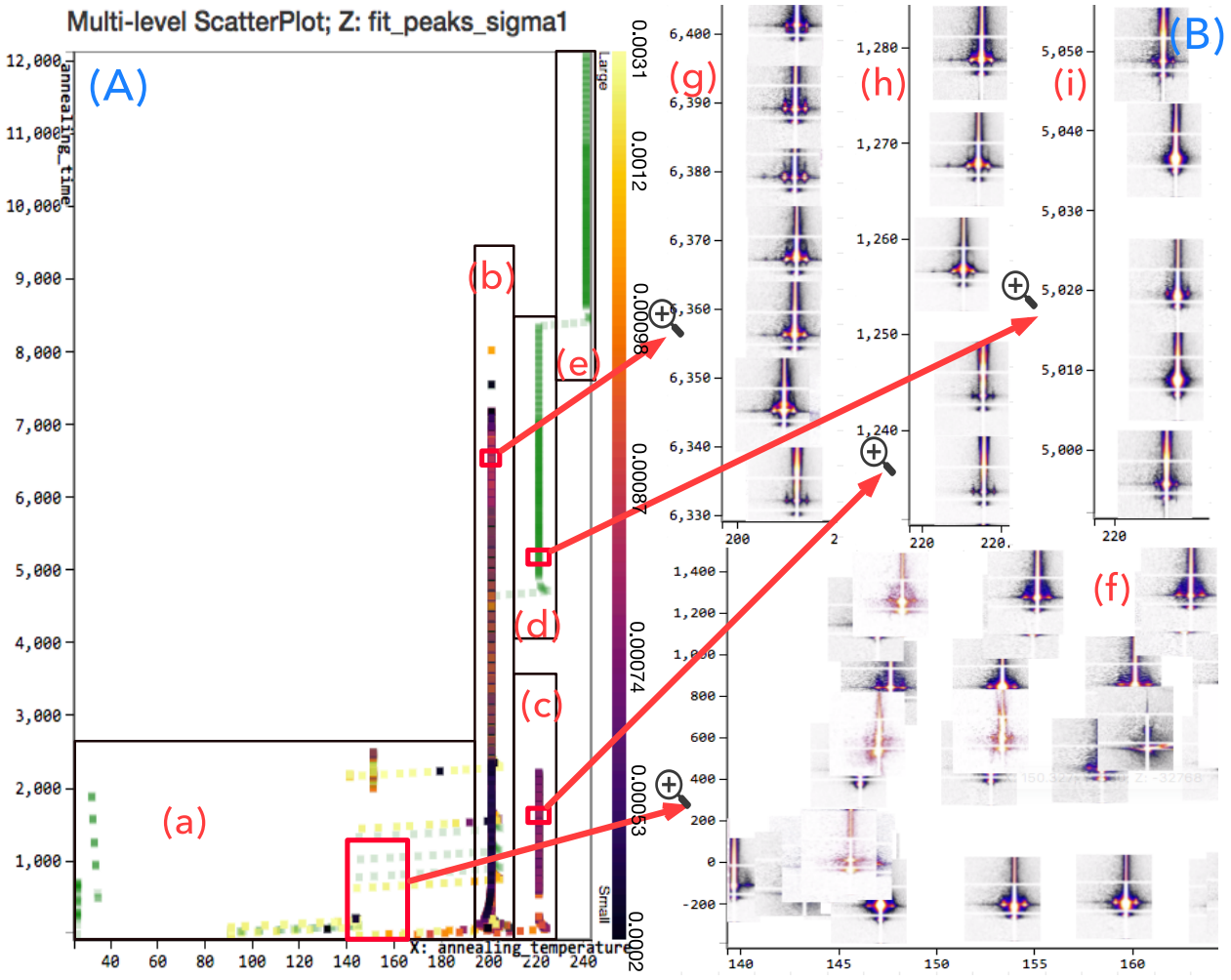


Fig. 8. Multi-level scatterplot. (A) The scatterplot of *annealing_time*, *sample_name* and *fit_peaks_sigma1* attributes with regards to x, y and z axes respectively. The samples with green color mean they have not been analyzed. The z axis values are encoded using colors.

brushed second largest set of grain sizes in the distribution 10.b and noticed a smooth and continuous color transition in the scatterplot 10.d. These were thus identified as highly-ordered samples. Further zooming confirmed that these sample images 10.f are robust, exhibiting bright and sharp scattering peaks. By using the cross filter and the scatterplot view, the user was thus able to identify the experimental conditions that gave rise to the best sample ordering. Overall, the cross filter enables free exploration of the dataset, in order to identify possibly interesting trends in the data, and to subsequently test scientific hypotheses.

7.4. User Feedback

In general, the users (domain scientists) like our *MultiSciView* system and regard it as a breakthrough since there is no existing visualization approach optimized for the x-ray scattering experiment images. Before the utilization of our proposed system, domain experts reported to take much more efforts, but can only understand parts of material properties. For example, starting from an overview scatterplot, they could not conveniently zoom into details of a local image scattering pattern and zoom out later to check other locations, without multi-level scatterplot. To make comparisons of interesting samples, they rely on collecting images manually onto a dashboard. However, with the fast growing of scattering images, this process is very time consuming, and becomes more and more inefficient and unmanageable. Because of the ineffective visualizations, domain experts consider the exploration as validation processes of their hypotheses rather than new discovery. In addition, previously, the users were not able to identify attribute relations or find sample patterns directly.

Considering the tremendous visualization advancements of the MultiSciView system, the users enjoy using it a lot. One user spoke highly of multi-level scatterplot and attribute cross filters. He mentioned “multi-level scatterplot provides me a very easy way to explore the images. Previously, I need to put them in separate folders and manually open them to check one sample after another. When comparing pixel intensities, I need to open each tiff file one by one. This is very time consuming.” For the cross filter, the user described it as “powerful interactions” since it enables focus on a set of samples behavior. He was really surprised when finding the unexpected observation that “the larger the size of the blended material, the larger d_0 is”, using cross filters. He thought this is an important future re-

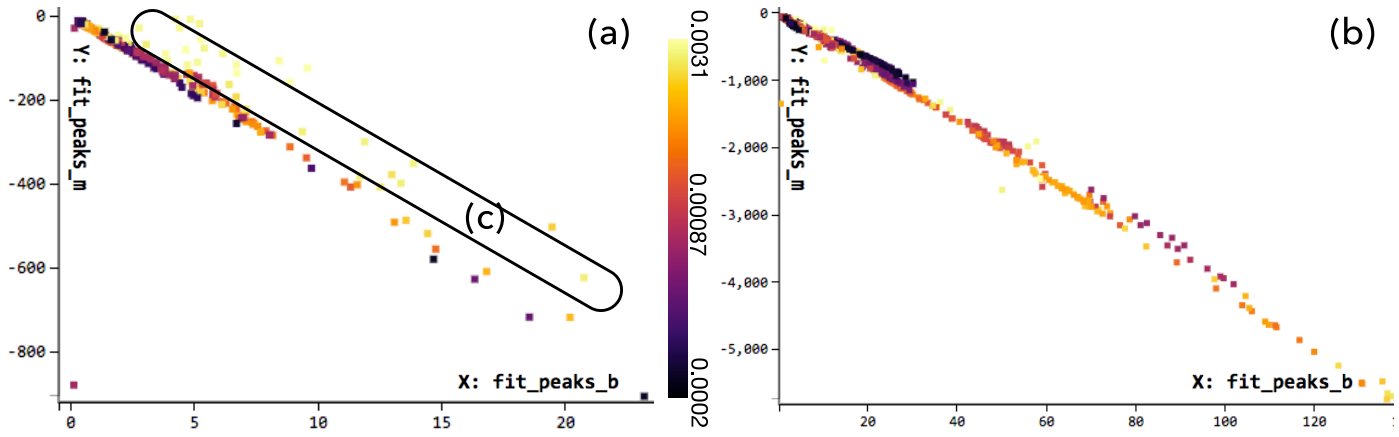


Fig. 9. Scatterplots whose x and y axes are denoted by *fit_peaks_b* and *fit_peaks_m* attributes respectively. (a) *C67* material. (b) *L74* material. The color legend (Z axis) is in the middle, encoding σ_1 attribute.

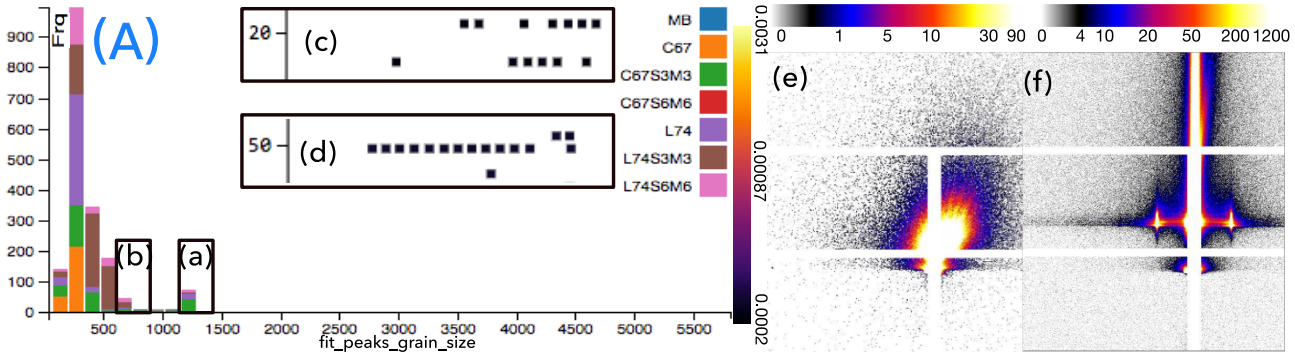


Fig. 10. (A) *grain_size* attribute stacked histogram based cross filter (e) Zoomed-in figure of (c) pattern. (f) Zoomed-in figure of (d) pattern.

search direction for synthesizing better quality materials. Due to the similarity of x-ray experiments, the users also recommend a potential application to other x-ray imaging modalities. This confirms the generalization of our system.

Besides the advantages, the users commented some limitations as well. The first one is attribute projection. He thought this visualization would be an effective way to understand all the attribute relations together. However, when observing the projection, he could only spot different attribute clusters but fail to interpret the inter-relations of these clusters. Currently, it can be rectified by comparing attributes from two clusters in the scatterplot visualization. But a more direct indication in the attribute project would be preferred. The second one is pixel value visualization. He conceded that it was great to zoom in for intensity checking. But he wondered whether the identification of peaks and largest intensity could be automatic. These two limitations will be our future improvement directions.

8. Conclusion

In this paper, we present *MultiSciView* a multivariate scientific image visualization and exploration system for x-ray scattering datasets. We devise three major visualizations to enable coordinated exploration across the image and attribute spaces. First, a scatterplot layout of the images exploration is presented which includes three levels of details: overview attribute level, image level, and pixel level. The attribute variation can thus be converted to a side-by-side image comparison, together with an auxiliary image detail visualization to provide further image evolution toward its actual pixel values. Cross filter visualization showing the attribute histograms and an attribute projection visualization are provided to accomplish the attribute space visualization and the coordinated exploration. Our case study demonstrates the benefits of our method in practical material science experiments.

As future works, we plan to further enhance our system based on user feedback. In details, we think proposing better and more interpretable distance metric and projection method for the attribute projection. Currently, we use the most common MDS embedding method, leading to neighbor distortion problem [18]. Then, peak fitting algorithm could be considered to automate peak recognition. Thus, the largest intensity within the peak region can then be encoded into the visualization. Moreover, we plan to refine the side-by-side image visualization to avoid image occlusion and further enhance the interaction among coordinated visualizations. Finally, we will extend the use cases to other x-ray scenarios to evaluate the generalization.

Acknowledgment

This research was supported by two Lab Directed Research and Development projects 16-041 and 17-029 of Brookhaven National Laboratory. Research used resources of the Center for Functional Nanomaterials, and the National Synchrotron Light Source II, which are U.S. DOE Office of Science Facilities, at Brookhaven National Laboratory under Contract No. DE-SC0012704. Experiments were performed at the Complex Materials Scattering (CMS, 11-BM) beamline at NSLS-II. Partial suport also came from NSF grant IIS 1527200.

References

[1] Asimov, D., 1985. The grand tour: A tool for viewing multidimensional data. SIAM Journal on Scientific and Statistical Computing 6, 128–143. URL: https://doi.org/10.1137/0906011, doi:10.1137/0906011.

[2] Blaas, J., Botha, C.P., Post, F.H., 2007. Interactive visualization of multi-field medical data using linked physical and feature-space views, in: Proceedings of the 9th Joint Eurographics / IEEE VGTC Conference on Visualization, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland. pp. 123–130. URL: http://dx.doi.org/10.2312/VisSym/EuroVis07/123-130, doi:10.2312/VisSym/EuroVis07/123-130.

[3] Bonnet, N., Herbin, M., Vautrot, P., 1997. Multivariate image analysis and segmentation in microanalysis. Scanning Microsc 11, 1–21.

[4] Castelletto, V., Hamley, I.W., 2004. Morphologies of block copolymer melts. Current Opinion in Solid State and Materials Science 8, 426–438.

[5] Chan, Y.H., Correa, C.D., Ma, K.L., 2010. Flow-based scatterplots for sensitivity analysis, in: 2010 IEEE Symposium on Visual Analytics Science and Technology, pp. 43–50. doi:10.1109/VAST.2010.5652460.

[6] Cleveland, W.S., McGill, M.E., 1988. Dynamic graphics for statistics. Statistics/Probability Series .

[7] Doerk, G.S., Yager, K.G., 2017a. Beyond native block copolymer morphologies. Molecular Systems Design & Engineering .

[8] Doerk, G.S., Yager, K.G., 2017b. Rapid ordering in wet brush block copolymer/homopolymer ternary blends. ACS nano .

[9] Doleisch, H., Gasser, M., Hauser, H., 2003. Interactive feature specification for focus+context visualization of complex simulation data, in: Proceedings of the Symposium on Data Visualisation 2003, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland. pp. 239–248. URL: http://dl.acm.org/citation.cfm?id=769922.769949.

[10] Elmqvist, N., Dragicevic, P., Fekete, J.D., 2008. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. IEEE Transactions on Visualization and Computer Graphics 14, 1539–1148. doi:10.1109/TVCG.2008.153.

[11] Fekete, J.D., Wijk, J.J., Stasko, J.T., North, C., 2008. Information visualization, Springer-Verlag, Berlin, Heidelberg. chapter The Value of Information Visualization, pp. 1–18. URL: http://dx.doi.org/10.1007/978-3-540-70956-5_1, doi:10.1007/978-3-540-70956-5_1.

[12] Gleicher, M., 2018. Considerations for visualizing comparison. IEEE Transactions on Visualization and Computer Graphics 24, 413–423. doi:10.1109/TVCG.2017.2744199.

[13] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C.D., Roberts, J.C., 2011. Visual comparison for information visualization. Information Visualization 10, 289–309. URL: https://doi.org/10.1177/1473871611416549, doi:10.1177/1473871611416549.

[14] Gresh, D.L., Rogowitz, B.E., Winslow, R.L., Scollan, D.F., Yung, C.K., 2000. Weave: a system for visually linking 3-d and statistical visualizations applied to cardiac simulation and measurement data, in: Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145), pp. 489–492. doi:10.1109/VISUAL.2000.885739.

[15] Jankun-Kelly, T.J., Ma, K.L., 2001. Visualization exploration and encapsulation via a spreadsheet-like interface. IEEE Transactions on Visualization and Computer Graphics 7, 275–287. URL: http://dx.doi.org/10.1109/2945.942695, doi:10.1109/2945.942695.

[16] Kehrer, J., Hauser, H., 2013. Visualization and visual analysis of multifaceted scientific data: A survey. IEEE Transactions on Visualization and Computer Graphics 19, 495–513.

[17] Keim, D.A., 2002. Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics 8, 1–8. doi:10.1109/2945.981847.

[18] Lespinats, S., Aupetit, M., 2011. Checkviz: Sanity check and topological clues for linear and non-linear mappings, in: Computer Graphics Forum, Wiley Online Library. pp. 113–125.

[19] Liu, S., Maljovec, D., Wang, B., Bremer, P.T., Pascucci, V., 2017. Visualizing high-dimensional data: Advances in the past decade. IEEE Transactions on Visualization and Computer Graphics 23, 1249–1268. doi:10.1109/TVCG.2016.2640960.

[20] Majewski, P.W., Yager, K.G., 2016. Rapid ordering of block copolymer thin films. Journal of Physics: Condensed Matter 28, 403002.

[21] Malik, M.M., Heinzl, C., Groeller, M.E., 2010. Comparative visualization for parameter studies of dataset series. IEEE Transactions on Visualization and Computer Graphics 16, 829–840. doi:10.1109/TVCG.2010.20.

[22] Müller-Buschbaum, P., 2016. Gisaxs and gisans as metrology technique for understanding the 3d morphology of block copolymer thin films. European Polymer Journal 81, 470–493.

[23] Rahman, A., Majewski, P.W., Doerk, G., Black, C.T., Yager, K.G., 2016. Non-native three-dimensional block copolymer morphologies. Nature communications 7, 13988.

[24] Roberts, J.C., 2007. State of the art: Coordinated multiple views in exploratory visualization, in: Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007), pp. 61–71. doi:10.1109/CMV.2007.20.

[25] Rbel, O., Weber, G.H., Kernen, S.V.E., Fowlkes, C.C., Hendriks, C.L.L., Simirenko, L., Shah, N.Y., Eisen, M.B., Biggin, M.D., Hagen, H., Sudar, D., Malik, J., Knowles, D.W., Hamann, B., 2006. Pointcloudxplore: Visual analysis of 3d gene expression data using physical views and parallel coordinates, in: Eurographics/IEEE-VGTC Symposium on Visualization Proceedings, pp. 203–210.

[26] Spenke, M., Beilken, C., 2000. Infozoom - analysing formula one racing results with an interactive data mining and visualisation tool, in: in Ebecken, N. Data mining II, pp. 455–464.

[27] Spenke, M., Beilken, C., Berlage, T., 1996. Focus: The interactive table for product comparison and selection, in: Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology, ACM, New York, NY, USA. pp. 41–50. URL: http://doi.acm.org/10.1145/237091.237097, doi:10.1145/237091.237097.

[28] Swayne, D.F., Lang, D.T., Buja, A., Cook, D., 2003. Ggobi: Evolving from xgobi into an extensible framework for interactive data visualization. Comput. Stat. Data Anal. 43, 423–444. URL: https://doi.org/10.1016/S0167-9473(02)00286-4, doi:10.1016/S0167-9473(02)00286-4.

[29] Ward, M.O., 1994. Xmdvtool: integrating multiple methods for visualizing multivariate data, in: Visualization, 1994., Visualization '94, Proceedings., IEEE Conference on, pp. 326–333. doi:10.1109/VISUAL.1994.346302.

[30] Yager, K.G., Lai, E., Black, C.T., 2014. Self-assembled phases of block copolymer blend thin films. ACS nano 8, 10582–10588.

[31] Yee, K.P., Swearingen, K., Li, K., Hearst, M., 2003. Faceted metadata for image search and browsing, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 401–408. URL: http://doi.acm.org/10.1145/642611.642681, doi:10.1145/642611.642681.

[32] Zhang, Z., McDonnell, K.T., Zadok, E., Mueller, K., 2015. Visual correlation analysis of numerical and categorical data on the correlation map. IEEE Transactions on Visualization and Computer Graphics 21, 289–303. URL: doi.ieeecomputersociety.org/10.1109/TVCG.2014.2350494, doi:10.1109/TVCG.2014.2350494.