Eurographics Conference on Visualization (EuroVis) (2012)
M. Meyer and T. Weinkauf (Editors)

CrystalExplorer: An Interactive Knowledge-Assisted System for Visual Design of Solar Cell Crystal Structures

Amal Aboulhassan, Ruipeng Li, Christopher Knox, Aram Amassian, and Markus Hadwiger

King Abdullah University of Science and Technology, Saudi Arabia

Abstract

Crystallography is a key research tool in materials science. The chemical properties of materials are often controlled by the geometric properties of crystals. Accordingly, visualizing the 3D structure of crystals is an important task in materials exploration. The current crystallography visualization systems are limited by focusing on the visualization of pre-defined crystal structures, and a lack of capabilities for editing and exploring different variations and levels of abstraction. To remove this limitation, we propose a new paradigm for knowledge-assisted visual exploration of crystals where the user can use semantic rules to define clusters of atoms with certain geometric properties. To test the usefulness of this system, we have applied it for the design of materials for solar cells. Using our proposed system, materials scientists can interactively create and visualize structures of interest inside the crystals in a relatively short time. This could not be achieved using their previous visualization workflow.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—

1. Introduction

Crystallography is a very important part of research in materials science. Nowadays, materials design is a critical aspect in many areas such as the manufacturing of transistors or solar cells. One of the main principal questions regarding the efficient design of such materials is about the relationship between the geometric properties and the chemical properties of a molecular structure.

Crystals are a set of molecules packed symmetrically in a lattice under what are known as Space Group operations [H02]. Many tools have been introduced for the visualization of direct geometric properties of atomic 3D structures of crystals, such as atomic positions and molecular topologies. These visualization techniques are mainly based on the classic balls-and-sticks model. Unfortunately, this model is not sufficient in many applications since the indirect geometric properties of atomic clusters need to be considered as well. Due to the amount of visual clutter inherent to balls-and-sticks visualizations, it is almost impossible to perceive individual clusters of atoms with certain geometric patterns, especially in big crystal structures.

In this paper, we propose a new paradigm for knowledge-assisted visual exploration of crystals. The goal of our system is to allow users to interactively create abstract geometry on top of detailed atomic 3D structures. These structures then allow users to focus on the regions of interest, and hide all other unnecessary details.

We propose a new variation of the conventional hierarchical oriented bounding box data structure [GL96], called Crystal Oriented Bounding Box Hierarchy (C-OBBH). This data structure enables hierarchical exploration of different geometric properties of clusters across the crystal and switch among them using different levels of detail.

In order to incorporate user knowledge into the process of building a C-OBBH, we propose a semantics-based approach that allows the user to feed the system with high-level semantic parameters. A set of pre-defined rules are used to translate the user-defined semantic parameters into geometric properties. These geometric properties are then used to hierarchically build the C-OBBH in a bottom-up fashion.

Our system provides two main contributions:

- Interactive knowledge-based clustering of crystal data.
- Visual encoding of geometric properties of atomic clusters inside crystals.

DOI: 10.2312/PE/EuroVisShort/EuroVisShort2012/031-035

The usefulness of the system is tested via a case study in cooperation with a materials science research group studying the design of photocells. We asked them to perform their primary research tasks with our system, and compared this process to their traditional workflow. By using our proposed system, the time needed by the user to perform these tasks is reduced from several hours to only minutes. Moreover, our system helped them to interactively select and visualize different subsets of the crystal corresponding to specific semantic goals. This advanced feature was not possible in their traditional workflow.

2. Previous Work

Current crystallography systems are mainly used to visualize symmetric properties such as the packing of molecules and the properties of symmetry axes and planes. Examples include CrystalMaker (http://www.crystalmaker.com/), Mercury (http://www.ccdc.cam.ac.uk/products/mercury/), and JMol [RB10], which is an OpenSource Java-based tool developed for educational purposes. While these systems are helpful in understanding crystal structures, they are of limited help in knowledge-based exploration where the user needs to create and visualize new structures of interest inside the crystal. We aim at removing this limitation by providing a system that allows the user to explore geometric properties of user-defined clusters inside the crystal and relate those to semantic parameters.

The idea of visualizing abstract geometric primitives for clusters of atoms has been first introduced in the field of biology and is known as Protein Secondary Structures [C83]. These are specific clusters of atoms of certain geometric properties that are of concern to biologists. We use a similar concept in this work, but instead of mapping a pre-defined set of primitives, we allow the user to explore semantic geometric primitives interactively. These semantic parameters are mapped to data clusters using structure analysis concepts such as [AF06] and collision detection literature [GL96].

Other techniques have been introduced for the visualization of crystals, such as [MH04] and [MJ06]. On the other hand, many methodologies have been recently introduced that handle interactive molecular visualization, e.g., [LB10], [JV09], [BD04], and [GR07]. The main contribution of the presented work is that it incorporates semantic, knowledge-based techniques into the exploration of crystal molecules.

3. System Design

The electron distribution around atoms is the core factor in determining most of the chemical properties of molecules. Since there is a strong relation between the molecules' spatial geometric properties and electron motion, chemists make great use of modeling and visualizing this data. In our proposed system, we allow material designers to interactively build clusters of atoms of certain geometric semantics that define a certain electron motion pattern, and thus a general chemical property such as electrical charge transfer.

3.1. Crystal Oriented Bounding Box Hierarchy (C-OBBH)

We propose a new data structure that we call Crystal Oriented bounding Box Hierarchy (C-OBBH). C-OBBH hierarchically stores geometric primitives built through interactive clustering of molecular 3D structures.

We build C-OBBH on the concept of the conventional oriented bounding box tree [GL96]. The difference is that C-OBBH is not a simple tree with uniform nodes. Rather, it is a set of distinct conceptual levels of abstraction that include *molecule*, *unit cell*, *crystal*, and *group of crystals*; from the lowest to the highest level of abstraction. This conceptual hierarchy matches the intuitive interpretation of crystallographers for crystals, and therefore suits our knowledge-assisted clustering technique. Once the user performs some clustering at a certain level, the system automatically broadcasts it to the corresponding symmetrical positions according to the symmetry operators defined in the input crystal data. The result is a 3D conceptual hierarchical representation of the crystal clusters. This can help the user to switch between different levels of abstraction and explore the corresponding geometric properties accordingly.

3.2. Knowledge-Assisted C-OBBH Editing

The user interactively builds a C-OBBH on top of the conventional balls-and-sticks visual representation of crystals. The balls-and-sticks diagram allows the user to view the molecules in a 3D graph representation and to select atoms interactively. This set of selected atoms is then used to define initial geometric primitives in the lowest level of abstraction (*molecule*). As the scope of the crystal exploration becomes larger, automatic operations are used instead.

Initially, the user is allowed to interactively define the set P: the set of initial primitives at the molecule level. We provide a set of graph editing operations (including split, merge, add, and delete) to match the intuitive graph nature of molecules. Hence, with a few mouse clicks, the user can produce the required clustering. At the higher levels of C-OBBH, the user defines criteria for merging clusters using semantic rules. The input to this stage is P. The geometric aspects in our current study include the following: shape, collision, distance, and style. The program maps these semantic rules to clustering operations using a set of pre-defined cost functions, as shown in Figure 1.

If (Min1 <= C-Shape <= Max1) **and** (Min2 <= C-Collision <= Max2) **and**
(Min3 <= C-Distance <= Max3) **and** (C-Style = Style i)
Then
(Merge/Don't Merge) **and** (Apply Style j)

Figure 1: *General kernel of a rule that maps semantics to clustering operations.*

We design a set of cost functions to cover the main geometric aspects as follows.

C-shape: This cost function measures the distance between the fitted primitive and the cluster points. The system computes a primitive fitting formula to find the parameters of the best abstract geometric shape fitting into the current cluster. Then the distance function is used to give more accurate information about whether this shape is deformed or not.

C-Collision: This cost function measures the amount of collision between two clusters. We use the potential of the oriented bounding box technique in collision detection to allow the user to visualize the amount of intersection between clusters both in 2D and 3D.

C-Distance: This cost function is used to measure distances between clusters and compare them to a user-defined atomic distance.

C-Style: This cost function is used to perform clustering according to some visual style (such as color, material, etc.). This way the user can specify a certain style for clusters. Moreover, he can also use styles as a rule for clustering as shown in Figure 1.

Once the cost functions are computed, thresholds are used to define semantic regions in each case.

4. Case Study

We have implemented a prototype and applied it in exploring crystalline materials for solar cell applications. Solar cell designers need to study candidate paths for charge transfer based on geometric criteria. For example, they believe that clusters of planar shapes with big overlapping areas are good charge paths, as opposed to non-planar clusters. Accordingly, they need to visualize the data and compare charge paths in different crystals, to enable making a decision about the material to use. In what follows, we demonstrate how users can interactively build such clusters using the design criteria we discussed earlier.

4.1. GUI

We have developed an Avizo (`http://www.vsg3d.com/avizo/overview`) plugin using C++. First, the user loads a crystal PdB file (`http://www.rcsb.org/pdb/`) into a balls-and-sticks diagram. Then, the user chooses a C-OBBH level and enters a combination of semantic parameters using a GUI. An example is shown in Figure 2.

4.2. Rules Execution

For the current case study, we present examples of the cost functions we have designed as follows.

- **Shape:** C-Shape is based on a plane-fitting for this specific application. This concept could be extended to incorporate other geometric primitives [AF06]. We use Principle Component Analysis (PCA) to compute the normal of the plane from the following covariance matrix.

Geometrical Aspect	Semantic Value	Threshold	
		Min	Max
Shape	Perfect Plane	0%	1%
	Deformed Plane	1%	5%
	Non-plane	5%	>5%
Distance	Small	0	x1
	Big	x1	>x1
Style 1	Color	Blue	Red
	Material	Opaque	
Default	None		

Figure 2: *Example semantics defined using the GUI and the corresponding internal thresholds.*

$$Cov = \sum_i (v_i - v)(v_i - v)^T \quad (1)$$

Where

$$v = \sum_i v_i / N, \quad (2)$$

is the set of N atom vertices belonging to the current cluster.
Finally, the normal n is the eigenvector corresponding to the minimum eigenvalue of the covariance matrix. Accordingly, C-Shape is computed as:

$$C_{shape} = \sum_i (n(v_i - v))^2 \quad (3)$$

We define how far a plane is from perfection as (C-shape /n)*100.
- **Distance:** Since the distance range is usually defined according to the atom type, the range is explicitly defined by the user. This distance is computed according to the shape type; e.g. shortest distance in the case of planes, or the distance between centroids otherwise.
- **Style:** Two styles are applied to the bounding boxes according to the semantics described in the table as shown in Figure 2.

Finally, the threshold value ranges are defined as shown in Figure 2. When the user finishes the selections, a sequence of rules can be fired by the system engine as shown in Figure 4.

5. Results and Discussion

The target group is a team of chemists and materials scientists. In our current case study, chemists depend mainly on manual calculations for exploring crystals, while material designers depend on conventional crystallography visualization software for exploring their crystal data. Accordingly, the variation in the team's specialities provided a good test

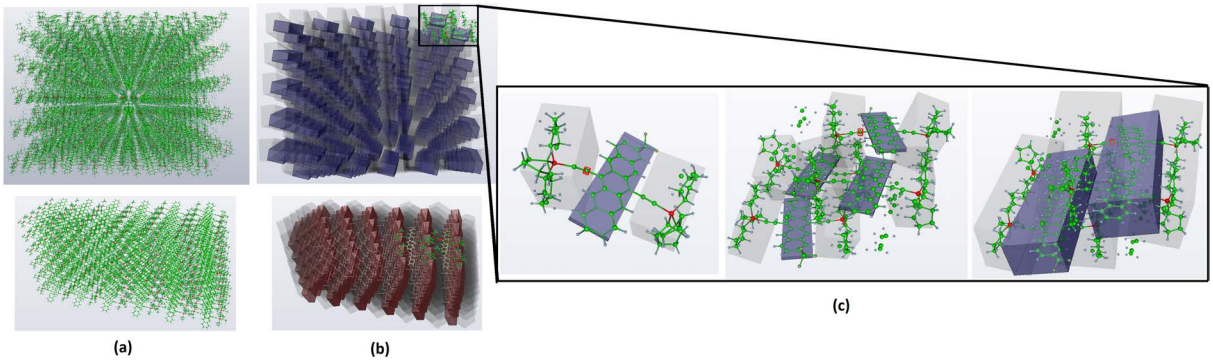

Figure 3: *Partial results produced interactively in our system. Column (a) shows the balls-and-sticks representation of two different crystal datasets. The visual clutter of this representation makes it hard to compare structures of interest between the two crystals. Column (b) shows the corresponding electrical charge paths created interactively. These structures are created by a user rule to group planar shapes within a certain distance. For example, the user can visually distinguish flat charge paths from sheared ones. In addition, he can determine the orientation of these paths within the whole crystal. Coloring the planes gives an indication whether the planes are deformed (blue) or not (red). (c) shows the electrical path in lower levels of detail (*molecule *and* unit cell*). At the molecule level, the user defined different clusters. Our system automatically detected sufficiently planar structures and applied the respective visual style specified by the user, as deformation is otherwise hard to perceive.*

environment for the usefulness of our proposed system. We have used the system to perform three main exploration tasks that the group performed and then compared against their previous workflow. We have also used samples of their data of medium size (5x5 unit cells; from around 10000 to 16000 atoms as shown in Figure 3).

- **Task 1: Visualizing good and bad charge paths.** The system allows the user to interactively create good and bad paths for charge transfer, depending on semantic geometric parameters fed to the program as shown in Figure 3. The user's previous way of doing this was using Maya to manually specify geometry for charge paths, because all of their other tools do not support this task at all. This has the obvious disadvantage of being time-consuming: it can take days to produce one visual representation using manual rendering. On the other hand, using our system users can produce such representations in a few minutes and edit them interactively.

Semantic Rule	**Mapping Rule**
In the **Molecule Level**: **Apply Style1** on **Planes**	**If** (0% <= C-Shape <= 5%) **Then** (Merge) **and** (Apply Style 1)
In the **Unit Cell level**:**Merge** clusters with **Small Distance** and **Style 1**.Then **apply style 1**	**If** (0.0 <= C-Distance <= X1) **and** (C-Style = Style 1) **Then** (Merge) **and** (Apply Style 1)

Figure 4: *Example rules mapping user-specified semantic instructions into clustering operations.*

- **Task 2: Visual encoding of the geometric information of atom clusters.** The users can visually represent information such as the amount of deformation in planes and relative distances between them, as shown in Figure 3. Using the previous workflow, they mainly depended on manual calculations or visual comparison using the highly cluttered balls-and-sticks diagram. Accordingly, they could not investigate information like relative distances and overlaps, which is now possible using our system.
- **Task 3: Comparing different geometric structures.** This can be achieved by using different visual encodings built interactively as shown in Figure 3.

6. Conclusions and Future Work

In this paper, we have proposed a system that allows materials scientists and chemists to create clusters of atoms from crystal data based on desired geometric properties. Our interaction paradigm also incorporates semantics to facilitate exploring big crystal structures interactively. This approach can be used in the design process, as well as for presentation purposes.

In the future, we want to investigate more advanced visualization techniques for enhancing the visual presentation of the abstracted crystal geometry. Moreover, we plan to widen the scope of the application to include more geometric aspects and primitives of crystallography in the field of materials science, as well as integrating the visualization and geometry editing with a simulation of the resulting material properties.

References

[RB10] RULE G., BAJZEK D. , AND KESSLER A.: Molecular visualization in STEM education: Leveraging Jmol in an integrated assessment platform. *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2010* , (2010), pp. 304–309. 2

[H02] HAHN T.: International tables for crystallography: Space-group symmetry. *International Union of Crystallography by Kluwer Academic Publishers*, (2002). 1

[GL96] GOTTSCHALK S., LIN M., AND MANOCHA D.: OBB-Tree: A hierarchical structure for rapid interference detection. *Proc. SIGGRAPH '96*, (1996), pp. 171–180. 1, 2

[AF06] ATTENE, M., FALCIDIENO, B., AND SPAGNUOLO, M.: Hierarchical mesh segmentation based on fitting primitives. *The Visual Computer: International Journal of Computer Graphics*, (2006), vol. 2, pp. 181–193. 2, 3

[C83] CONNOLLY M. L.: Analytical molecular surface calculation. *Journal of Applied Crystallography*, (1983), vol. 16, pp. 548– 558. 2

[MJ06] MEHTA K., AND JANKUN-KELLY T.J. : Detection and Visualization of Defects in 3D Unstructured Models of Nematic Liquid Crystals. *IEEE Transactions on Visualization and Computer Graphics*, (2006), vol. 12, pp. 1045– 1051. 2

[MH04] MEHTA S., HAZZARD K., AND MACHIRAJU R. : Detection and Visualization of Anomalous Structures in Molecular Dynamics Simulation Data. *IEEE Visualization*, (2004), pp. 465– 471. 2

[LB10] LINDOW N., BAUM D., PROHASKA S., AND HEGE H.-C.: Accelerated Visualization of Dynamic Molecular Surfaces. *Journal compilation l' 2010 The Eurographics Association and Blackwell Publishing Ltd.*, (2010) , vol. 29. 2

[JV09] JANG Y., VARETTO U., MA K., PFISTER H., MÖLLER T.: Interactive Volume Rendering of Functional Representations in Quantum Chemistry. *Visualization and Computer Graphics, IEEE Transactions on*, (2009), pp. 1579– 5186.. 2

[BD04] BAJAJ C., DJEU P., SIDDAVANAHALLI V., THANE A.: TexMol: Interactive Visual Exploration of Large Flexible Multi-Component Molecular Complexes. *In IEEE Visualization (VIS)*, (2004), pp. 243 –250. 2

[GR07] GROTTEL S., REINA G., VRABEC J., AND ERTL T. : Visual Verification and Analysis of Cluster Detection for Molecular Dynamics. *IEEE Transactions on Visualization and Computer Graphics*, (2007), vol. 13, pp. 1624 –1631. 2