A Computer-aided Design Tool with Semiautomatic Image-Processing Features for Visualizing Biological Pathways

Sung-Il Ham[2], San-Duk Yang[1], Chin-Ting Thong[1] and Hyun-Seok Park[1,2]*

[1]Department of Computer Science, Ewha Womans University, Seoul 120-750, Korea, [2]Department of Architectural Engineering, Yonsei University, Seoul 120-749, Korea

Abstract

The explosion in biological data resulting from high-throughput experiments requires new software tools to manipulate and display pathways in a way that can integrate disparate sources of information. A visual Java-based CAD tool for drawing and annotating biological pathways with semiautomatic image-processing features is described in this paper. The result of the image-editing process is an XML file for the appropriate links. This tool integrates the pathway images and XML file sources. The system has facilities for linking graphical objects to external databases and is capable of reproducing existing visual representations of pathway maps.

Availability: PPE has been implemented using Java and relies on Open Source Software; it includes some external image-processing modules, imported from ImageJ (http://rsb.info.nih.gov/ij/). The system is only a pre-alpha release and needs to be improved with a user interface to be fully automated. You may download a demo version of an executable file (ppe1.0_executable.jar_all_os.zip) from Sourceforge.net (http://www.sourceforge.net/projects/j2dpathway). However, some of the ImageJ modules are excluded in the demo version. Contact neo@ewha.ac.kr or archispace@yonsei.ac.kr if you are interested in the full version of the system or the source code for possible collaboration.

Keywords: bioinformatics, computer-aided design, metabolic pathway visualization, systems biology

*Corresponding author: E-mail neo@ewha.ac.kr
Tel +82-2-3277-2831, Fax +82-2-3277-2306

Overview of the System

CAD (Computer-Aided Design) has been traditionally used to assist in modeling in various fields, and advances in the biological community have created new uses for CAD with many novel biomedical applications (Demir *et al.*, 2002; Holford *et al.*, 2005; Sorokin *et al.*, 2006; Sun, 2005). Especially, CAD can be a visualization gateway to gather numerous biological pathway resources, not only from biological pathway databases but also from static images and research articles.
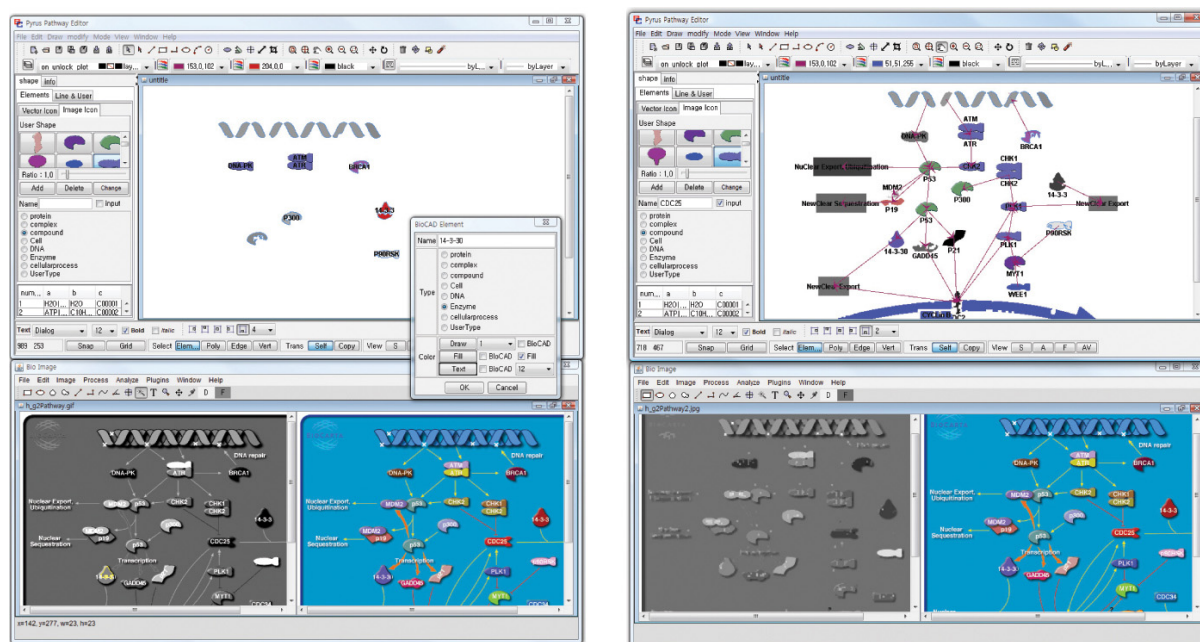
The PPE (Pyrus Pathway Editor) system, a Java-based visual tool for drawing and annotating biological pathways, is described in this paper. It is a graphical editor for biological pathways, and it integrates the pathway image and XML file source. Fig. 1 shows a snapshot of the PPE system.

The PPE window contains a menu bar, tool bar, status bar, and a progress bar at the top of the screen. Windows can be dragged around the screen and resized. The tool bar contains tools for making selections, for zooming and scrolling images, and for changing the drawing color. The status bar, when the cursor is over an image, displays pixel coordinates and values. The progress bar shows the progress of time-consuming operations. PPE provides several common visual notations but also allows the user to define his own notations. The shape menu in the left window, in
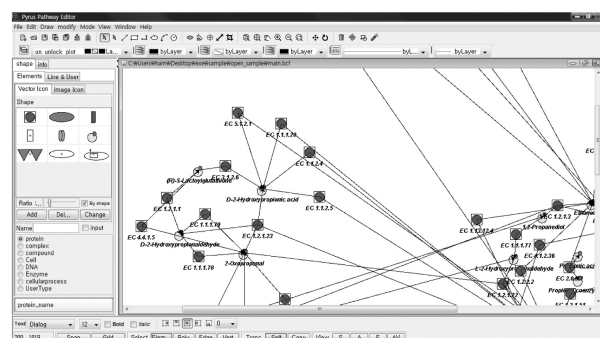
Fig. 1. Snapshot of the PPE System (Mode 1): PPE supports most standard image-processing functions, such as contrast manipulation and smoothing. It does geometric transformations, such as scaling and rotation. Images can be zoomed.

(a) User clicks on each node to set up the node type.   (b) User manually clicks on nodes and connects each node.

Fig. 2. Semiautomatic Image Processing of Biological Pathways (Mode 2): This exemplary image source is from BioCARTA (http://www.biocarta.com).

Fig. 1, can be used to select the shapes of the biological nodes. Users can annotate map elements with biological information or link them to external databases. The File/Import submenu provides access to plug-ins for reading XML files and for loading images. Thus, maps can be drawn in several modes.

First, they can be imported from external XML file sources and laid out in the editor window, as in Fig. 1. In the case that layout information is not provided in the XML file, an automatic layout module will be applied to display biological pathways.

Second, it allows a user to import external image sources, process the image, convert the processed image into an XML format, and finally generate maps in the editor window. The File/Open command opens TIFF, GIF, and JPEG images. There are some excellent image sources, ranging from image pathway databases, such as KEGG (http://www.kegg.org/) and BioCARTA (http://www.biocarta.com), to images in numerous biological textbooks, research papers, and websites. These external images are initially displayed in additional windows below the main editing window, as in Fig. 2a. They will be semiautomatically processed, converted into an XML format, and displayed in the main editing window, as in Fig. 2b. Each node and edge in the original image has to be confirmed by a user, one by one; double-clicking on any line or image nodes in the bottom window will result in the Object in the editor

window. By selecting an appropriate node type from the pop-up window, the user can create a new object in the main window. The types and shapes of nodes are predefined. The node types include protein-node, chemical compound-node, DNA-node, RNA-node, enzyme-node, and others. In addition, users can define their own node type. By selecting the exact concept from the pop-up menu, the user can specify an appropriate edge type, too.

Discussion and Future Direction

KEGG and BioCARTA visualize pathways in a static way. Pathway diagrams are manually drawn and stored as bitmap image files. While this visualization style offers good pathway representation, it does not provide the facilities to create and visualize dynamic pathways. Fortunately, KEGG also provides XML files. In the case that layout information is not provided in the XML file, an automatic layout module will be applied to display biological pathways. However, there are cases in which XML files are not provided in many other databases. In this case, we have to convert an image file directly to an XML file.

In this paper, a computer-aided approach to modify biological pathways, both in image and XML file format, is presented. To the best of our knowledge, we have not found any attempt to process biological pathway

(a) Original Image   (b) Filtered-out Image

Fig. 3. Filtering Out Characters in the Image. We borrowed part of the modules from ImageJ (Abramoff *et al.*, 2004).

images within the CAD environment. However, there are some thorny issues that need to be solved to fully automate this process. Especially, an optical character recognition module should be added in the future to recognize characters in the image. Currently, the characters in the original images are filtered out, and the names and types of each node are manually annotated by humans in a later process, as in Fig. 3. Once the characters can be recognized, the naming process of each node, by human intervention, can be fully automated in the majority of cases.

References

Abramoff, M.D., Magelhaes, P.J., and Ram, S.J. (2004). Image processing with image. *J. Biophotonics International* 11, 36-42.

Demir, E., Babur, O., Dogrusoz, U., Gursoy, A., Nisanci, G., Cetin-Atalay, R., and Ozturk, M. (2002). PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinform.* 18, 996-1003.

Holford, M., Li, N., Nadkarni, P., and Zhao, H. (2005). VitaPad: visualization tools for the analysis of pathway data. *Bioinform.* 21, 1596-1602.

Sorokin, A., Pally, K., Selkov, A., Derain, O.V., Dronov, S., Ghazal, P., and Goryanin, I. (2006). The pathway editor: a tool for managing complex biological networks. *IBM J. Res. Dev.* 50, 561-573.

Sun, W. (2005). Bio-CAD. *Computer-Aided Design* 37, 1095-1096.