Towards the Visualization of Software Systems as 3D Forests: the CodeTrees Environment

Ugo Erra and Giuseppe Scanniello
Dipartimento di Matematica e Informatica
Università della Basilicata
Potenza, Italy
{ugo.erra, giuseppe.scanniello}@unibas.it

ABSTRACT

We present an approach based on a forest metaphor to ease the comprehension of object oriented software systems. Software systems are represented as forests of trees that users can navigate and interact with. We also describe here the mapping of the information of the source code in meaningful ways to take advantages of familiar concepts such as agglomerates of trees (or sub-forest), trunk, branches, leaves, and color of the leaves. The approach has been implemented in a prototype of a 3D environment, namely CodeTrees. To assess the validity of the approach and environment, we have also conducted a preliminary empirical evaluation on three open source software systems implemented in the programming languages Java and C++.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: Graphical User Interface; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques

Keywords

Software visualization, Software maintenance, forest metaphor

1. INTRODUCTION

Software maintenance is essential in the evolution of software systems and represents one of the most expensive, time consuming, and challenging phase of the development process. Maintenance starts after the delivery of the first version of the system and lasts much longer than the initial development process [1]. As shown in the survey by Erlikh [2], the costs needed to perform maintenance operations range from 85% to 90% of the total cost of a software project.

During the maintenance phase, a software system is continuously changed and enhanced because of the execution of maintenance operations that are carried out for several reasons (e.g., to correct faults or to improve quality requirements) [3]. Whatever is the maintenance operation, a maintainer has to comprehend source code implemented by others [4]. Therefore, it is easy to imagine that the greater part of the cost and effort for accomplishing maintenance tasks is due to the comprehension of the source code. There are several reasons that make the comprehension even more costly and complex, they range from the size of the software to its overall quality.

Software visualization [5], [6] has been extensively and successfully explored and used in the software maintenance and program comprehension, in particular. Researchers have proposed metaphors and supporting tools based on 2D and 3D techniques [7], [8], [9]. The proposed metaphors and tools often fail to show significant information of a software system to improve its comprehension (e.g., public attributes and methods).

In this paper, we propose a 3D visualization metaphor for depicting object oriented software system and a supporting tool, named CodeTrees. In particular, a given software is visualized as a forest of trees that users (also maintainers in the following) can navigate and interact with. Visual properties of trees (e.g., trunks, branches, and leaves) are mapped according to well defined rules with the metrics extracted from source code.

The metaphor is new from the point of view of software visualization and, compared to previous metaphors, provides a fine-grained representation of the entire software system and a large-grained representation of the classes and the packages they reside in. Therefore, we display classes as trees and packages as agglomerates of trees (from here on, sub-forests or simply forests). The shape (trunk and foliage) of each tree is characterized by source code metrics of the class it represents.

To validate the metaphor and CodeTrees, we have also conducted a preliminary case study on three well known open source software systems implemented in Java and C++: JEdit, ArtOfIllusion, and FileZilla.

The main contributions of the paper can be summarized as follows:

- A visualization metaphor based on forests of trees;
- A prototype of software tool (i.e., CodeTrees) implementing the metaphor;
- A preliminary case study on small/medium open source software systems implemented in Java and C++.

1Please read the paper on-screen or as a color-printed paper version, we make extensive use of color pictures.
The remainder of the paper is organized as follows: in Section 2, we review previous software visualization metaphors based on natural environments (e.g., the solar system) or real objects (e.g., a jigsaw puzzle). In Section 3, we describe the proposed metaphor. In Section 4, we present the preliminary results of the conducted empirical evaluation. Finally, Section 5 concludes and discusses possible future directions for our research.

2. RELATED WORK

In the literature there are some metaphors to visually represent software systems as natural environments. Among these, the city metaphor is one of the most explored (e.g., [10], [11], [12], [13]). For example, Wettel and Lanza [14] propose a city metaphor for the comprehension of object-oriented software systems. Similar to our proposal, the metaphor proposes a large and low-scale understanding of software. Classes are represented as buildings and packages as districts. The metaphor is implemented in the CodeCity tool and to improve the realistic aspect of the city, authors focused on the design of urban domain. Our metaphor is different because the fine-grained representation of classes takes into account a larger number of metrics (e.g., number of public methods). However, we plan users’ study to understand whether the use of more information results in a deeper understanding of the system.

In a recent paper, the same authors [15] present a controlled experiment with professionals to assess the validity of their metaphor and supporting tool. The results show that the CodeCity tool leads to a statistically significant improvement in terms of task correctness. The results also indicate that the use of the tool statistically decreases the task completion time.

Graham et al. [16] propose a solar system metaphor where suns represent a package, planets are classes, and orbits represent the inheritance level within the package. Such metaphor is used as a means to analyze either static or evolving code to perceive in real time suspected areas of risk within the code base.

Martinez et al. [17] suggest a metaphor based on landscape, whose main objective is to visualize the integrated representation of software development processes. The metaphor is conceived to describe several aspects of development process. Differently from our proposal, the metaphor does not provide means to visualize source code.

Ghandar et al. propose a jigsaw puzzle metaphor in [18], where each component of a system is modeled as a piece of a jigsaw puzzle. The most remarkable characteristic of the metaphor concerns the visualization of the system complexity that is represented as a pattern on the surface of the piece. The metaphor does not provide a view at class granularity level. This is one of the most remarkable differences with respect to our proposal.

A botanical tree metaphor is proposed in [19]. Differently from us, the authors suggest forests of trees for the visualization of huge hierarchical structures and apply the proposed metaphor to the visualization of directory structures. Directories, files, and their relations are visualized using trees. The approach is basically a natural visual metaphor for information hierarchically structured. Several are the differences with respect to our proposal. For example, we propose here a metaphor for the visualization of software systems, while in [19] the authors propose trees for depicting hierarchical structures in file systems. Further, we use trees as fine-grained representation of classes.

Several are the differences with respect to the approaches discussed above. The most remarkable one is that our proposal offers a proper representation for methods, attributes, and source code content. A possible drawback that affects our metaphor is that it could result complex in case the maintainer is not properly trained. This issue is directly connected to the considerable amount of information our metaphor is able to visually summarize.

3. THE METAPHOR

We propose a 3D visualization metaphor that depicts object-oriented software systems as forests of trees. Maintainers can freely navigate and interact with the forests to improve the comprehension of the systems. The reasons for defining this metaphor can be summarized as follows:

- A tree has a familiar shape. We use trees to visually summarize complex information in a natural way.
- A tree is a complex structure composed of unmistakable elements such as trunk, branches, leaves, and color. Several aspects of a software system can be represented through the visual properties of a tree.
- The class proliferation problem is a major cause of failure in object-oriented development. A forest of trees provides developers with a large-scale understanding of the design and the proliferation of classes avoiding exponential explosion in the number of classes.

Our main goal is a large-scale understanding of a software system because the overall result is a visualization of all the contained classes. The metaphor also provides a fine-grained representation of the components of a software in terms of individual classes. This is because a visualization of each class is possible zooming on the corresponding tree.

3.1 The Rendering of the Trees

The model upon which we based the creation and the rendering of the trees is based on the Weber and Penn approach [20]. This approach uses an intuitive model to design the geometrical structure of trees. It handles properties/parameters to modify the shape of a tree. There are general parameters to control: the height of the tree, the width of the trunk, the level of recursion of the branches, and so on. Further, there are also parameters that control more specific aspects of a tree (e.g., leaves orientation).

The approach needs no knowledge of botany and complex mathematical principles. For example, each branch may have similarity with its parent and inherit attributes from it. Due to this concern, all the branches are influenced by the primary branches that mainly depend from the tree height. In our case, this is not an actual drawback because a primary aim of the Weber and Penn approach is to get a plausible result with a few parameters. However, we consider a subset of the parameters needed to render a tree, thus affecting its shape (trunk and foliage) according to the main characteristics of the class the tree represents. The interested reader can read the paper by Weber and Penn [20] to get a deep description of the algorithm and to know how each parameter affects the creation and the shape of a tree.
3.2 Mapping Between Tree Properties and Source Metrics

We identified the following set of visual properties of trees that can give information about the system to analyze (or to comprehend in case of maintenance tasks): height, branch number, branch direction, leaf number, leaf color, leaves size, and base size (i.e., the part of trunk without branches).

The considered metrics used to influence the appearance of a tree are: lines of code (LOCs), lines of comment (CLOCs), number of attributes (NOAs), number of public methods (NEMs), number of private methods (NOMs), and the total number of methods (NEOMs). Although there many other design metrics available in the literature [21], we considered only the most widely known (e.g., [21]). The rationale for this choice relies on the fact that more complex and badly known metrics may complicate the metaphor. However, the use of different metrics is subject of future work.

The mappings between the properties of a tree and the selected metrics were based on the experience we gained in the program comprehension, software visualization, and development. We tried as much as possible to find intuitive mappings to ease the comprehension of a class.

Similarly to [14], we mapped the height of a tree with the LOCs metric to denote the size of a class (without take into account CLOCs). The number of branches that sprout from the trunk corresponds to NEOMs. A tree with few branches represents a class with few methods. To highlight the number of public methods we use branch orientation. In case the value NEMs/NEOMs is close to one, the class has many public methods and its tree has branches pointing out. Conversely, if NEMs/NEOMs tends to zero, the class has many private methods and its tree has branches oriented parallel to ground. We mapped the total number of leaves with NOAs and the size of the leaves results from 1/NOAs. In such way, a class with few attributes is represented as a tree with large leaves, so resulting more visible. While, a dense tree will represent a class with many attributes. We map the color of leaves as the ratio between the LOCs and the CLOCs. The color ranges from the green to brown. Thus, a class with lines of comments greater than lines of code is represented as a tree with a green foliage, brown otherwise. Finally, to visualize the amount of private methods we used the ratio NOMs/NEOMs that is visually represented as the base size of a tree. Then, if the value NOMs/NEOMs is close to one, the class has many private methods and the tree has a large base size. Conversely, if NOMs/NEOMs tends to zero, the class has many public methods and its tree has a small base size.

In Table 1, we summarize the mapping between metrics and visual properties. It is worth mentioning that some visual properties highlight the local characteristic of a given class (e.g., NOMs/NEOMs), while others make sense only in case trees are analyzed together (e.g., LOCs).

To illustrate how the mapping changes the appearance of a tree, we illustrate some sample classes. Figure 1 shows a tree of a possible class where the number of line of codes is high (see the foliage), the number of lines of comment is high, there are a considerable number of methods both public and private, but public methods are prevalent. Figure 4 shows a class with a high number of public methods and few attributes. This mostly happens when we are dealing with façade patterns. Finally, Figure 5 depicts a tree of a class with a huge number of methods. Many of these methods are public. The class also contains a large number of lines of comment.

The metrics extracted from the sample classes are shown in Table 2. Although the variability of the values for each metric, the tree of of each class maintains a good visual plausibility and naturalism.

3.3 A Forest of Trees

We represent a package as a sub-forest of the whole forest of trees. All the trees which represent classes inside a package are placed in a land square whose size is function of number of trees. To provide visibility to the smaller trees, we place them in front of the highest trees using a spiral pattern. In this way, highest trees will be placed in the cen-
Figure 3: FileZilla 3.0::sshbn. The tree represents a class with: a high number of lines of code, a high number of lines of comment, a number of public methods greater than private methods.

Figure 4: ArtOfIllusion 2.4.1::CustomDistortionTrack. The tree represents a class with: a large number of public methods, few private methods, and few attributes.

Figure 5: jEdit 4.2::jEdit. The tree represents a class with: a high number of public methods greater than private methods, a large number of lines of comment, a large number of attributes, and few private methods.

Table 1: Metrics and visual properties mapping.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Visual Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCs</td>
<td>Tall</td>
</tr>
<tr>
<td>NEOMs</td>
<td>Branches</td>
</tr>
<tr>
<td>NOAs</td>
<td>Number of leaves</td>
</tr>
<tr>
<td>NEMs/NEOMs</td>
<td>Branches orientation</td>
</tr>
<tr>
<td>1/NOAs</td>
<td>Size of leaves</td>
</tr>
<tr>
<td>LOCs/CLOCs</td>
<td>Color of the leaves (from green to brown)</td>
</tr>
<tr>
<td>NOMs/NEOMs</td>
<td>Base size</td>
</tr>
</tbody>
</table>

LOCs = Lines of Code  
CLOCs = Lines of Comment  
NOAs = Number of Attributes  
NEMs = Total Number of Methods  
NOMs = Number of Public Methods  
NEOMs = Number of Private Methods

<table>
<thead>
<tr>
<th>JAR</th>
<th>LOCs</th>
<th>CLOCs</th>
<th>NOAs</th>
<th>NEMs</th>
<th>NOMs</th>
<th>NEOMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>JarBundler</td>
<td>713</td>
<td>465</td>
<td>35</td>
<td>50</td>
<td>22</td>
<td>72</td>
</tr>
<tr>
<td>BrowserIORequest</td>
<td>248</td>
<td>71</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>sshbn</td>
<td>732</td>
<td>246</td>
<td>75</td>
<td>3</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>CustomDistortionTrack</td>
<td>362</td>
<td>51</td>
<td>0</td>
<td>34</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>jEdit</td>
<td>2488</td>
<td>1043</td>
<td>28</td>
<td>94</td>
<td>24</td>
<td>118</td>
</tr>
</tbody>
</table>

Table 2: Metrics of the sample classes.

An example of forest as depicted within CodeTrees is shown in Figure 6. The forest represents a software system composed of 7 packages. Some trees have no leaves, thus indicating that no attributes are present. Some trees have the branches that point out. When trees do not have leaves and the branches point out it is possible that we are in the case of a façade pattern. The leaves of some trees are brown, while others are green. In the first case, we can deduce that these classes are not properly commented, while the source code is commented in the latter case. The forest also contains a few trees whose branches are parallel to the ground. These trees have brown leaves, thus suggesting that the classes have a large number of private methods and are scarcely commented. Therefore, we can deduce that the greater part of the functionality implemented by these classes is only accessible through a few public methods.

3.4 Navigating within a Forest

The forest is visualized in real-time and the maintainer can navigate around the forest using a free-fly 3D camera. The maintainer can move inside the forest without limited movement capabilities and then it is possible also to pass through trees. The maintainer can then enlarges the tree...
of interest. The name of the class is also associated to the corresponding tree and visualized if required.

3.5 CodeTrees Implementation

The metaphor has been implemented in a prototype of a supporting system, named CodeTrees. The prototype is composed of three main software components. The former extracts the metrics for a given software system and then produces an XML file with all the information needed for the visualization. The design rationale for this choice relies on the fact that the extraction of the measures can be performed once for all. Further, this makes independent the extraction of the measures from the rendering of a forest and its trees. This component is implemented in Java.

The second component is aimed at graphically maps the tree parameters defined in the algorithm proposed in [20] and the extracted metrics. The component is implemented in Java and produces an XML file that is then used to make the rendering of a forest by our 3D engine.

To implement the 3D engine, we used the OpenTree library. This library provides 3D tree generation for real time applications such as games and visualization software. OpenTree is a cross-platform, engine-independent library written in C++. It implements the tree generation algorithm described in [20]. The library uses an array of vertex to generate mesh data necessary to render trees. Vertex can be used by any graphics library. We used in our implementation the OpenGL [22] graphics library.

4. EVALUATION

To validate CodeTrees and the underlying metaphor, we conducted a preliminarily case study on three object oriented software systems: JEdit, Art of Illusion, and FileZilla. The former two systems were implemented in Java, while the third in C++. We selected these systems: (i) to verify whether the validity of our proposal is affected by the programming language; (ii) because they have been widely used in the past to assess the validity of tools for supporting maintenance tasks. Another motivation for selecting these systems relies on the fact that they are very different in terms of implemented functionality:

- **JEdit** is a programmer’s text editor with an extensible plug-in architectures;
- **Art of Illusion** is a 3D modeling and rendering software system;
- **FileZilla** is a cross platform client for FTP, FTPS, and SFTP.

Some statistics of these systems are reported in Table 3. In particular, the table shows the numbers of classes and packages for JEdit, Art of Illusion, and FileZilla. For each system, the number of line of code and comment is shown as well. The latter rows report the number of methods (total, public, and private), the number of attributes, and the programming language used to implement each system, and the analyzed version.

![Figure 6: A Sample Forest composed of 7 packages.](image)

Table 3: Metrics of the three software systems.

<table>
<thead>
<tr>
<th></th>
<th>JEdit</th>
<th>Art of Illusion</th>
<th>FileZilla</th>
</tr>
</thead>
<tbody>
<tr>
<td># Classes</td>
<td>532</td>
<td>453</td>
<td>153</td>
</tr>
<tr>
<td># Packages</td>
<td>42</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>LOCs</td>
<td>102680</td>
<td>103404</td>
<td>71510</td>
</tr>
<tr>
<td>CLOCs</td>
<td>49743</td>
<td>18307</td>
<td>9976</td>
</tr>
<tr>
<td>NEMs</td>
<td>4920</td>
<td>5110</td>
<td>884</td>
</tr>
<tr>
<td>NOMs</td>
<td>1001</td>
<td>1010</td>
<td>332</td>
</tr>
<tr>
<td>NOAs</td>
<td>2994</td>
<td>2353</td>
<td>2299</td>
</tr>
<tr>
<td>Language</td>
<td>JAVA</td>
<td>JAVA</td>
<td>C++</td>
</tr>
<tr>
<td>Software Version</td>
<td>4.3</td>
<td>2.4.1</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Figure 7 shows the forests of the three software systems. On the top there is the JEdit forest, while in the middle the one for Art of Illusion. The forest of FileZilla is on the bottom. In the following, we discuss the findings achieved by applying the metaphor and the tool only from a quantitative perspective. In fact, the metaphor is based on the properties of trees and the metrics for each class in the forest. A qualitative analysis of the results is subject of future work.
Figure 7: From up to bottom the forest of JEdit, Art of Illusion, and FileZilla.
The big picture of the three forests shows that among the tree systems FileZilla has a smaller number of lines of comments. Further, this system contains the lowest number of classes and the majority of these classes contain a low number of lines of code. Further, the trees are sparse (i.e., subforests are not present) since the classes are not organized in packages being this system implemented in C++.

With regards to JEdit, we can observe that there are some classes that contained a high number of lines of comments. Among these classes, there are some classes with a larger number of public methods and with a few attributes. On the other hand, there are few classes with a large number of attributes and a few number of public methods. In the first case, the classes provide services to other classes and to make easy how to use these services the developers commented the code. In the latter case, the trees represent property classes.

The forest of Art of Illusion contain a big class with a large number of lines of code, attributes, and methods. A few lines of comment are present. Each sub-forest contains a tree without leaves and with branches that point out.

4.1 Further Results

The use of CodeTrees on the systems JEdit, Art of Illusion, and FileZilla leads also to the following considerations:

- **Scalability**. CodeTrees scales up well in terms of the size of the software systems to be visualized. However, for very large software systems (e.g., Eclipse 3.0) the tool slowed down, thus affecting the interactivity and the navigability of the forest. Several directions for our future work have been planned to improve scalability. For example, we will implement a GPU (Graphics Processing Unit) based version of CodeTrees.

- **Navigation and Interactivity**. Similarly to many 3D environments [14], CodeTrees allows the maintainer to move back or forward, hover left or right, orbit around the city, and change altitude. CodeTrees also enables the visualization of the name of each class through a label associated to the corresponding tree. A further feature to make the forest more realistic concerns the ground. In particular, the type (e.g., grass or stone) and the color (e.g., green or brown) of the ground can be changed if needed.

- **Completeness**. The defined metaphor and implemented 3D environment provide a fair amount of information for an overview of the system. They also offer a proper representation for methods, attributes, and comments. These characteristics make our approach different from the ones available in the literature (e.g., [14], [10]).

5. CONCLUSION AND FUTURE WORK

In this paper, we presented a 3D software visualization approach based on a forest of trees. The classes are represented as trees and the packages as sub-forests. The approach provides a general overview of the software giving a maintainer the possibility of carrying out a large-scale reasoning on it. Further, the classes within packages are represented as trees. The visual properties of each tree (e.g., trunks, branches, and leaves) are used to represent metrics of the class it represents. In this way, the maintainer has a large-grained representation of each class.

The metaphor has been implemented in prototype of a supporting tool, named CodeTrees. It provides features to navigate in the forest as a free fly 3D virtual camera. CodeTrees also implements zoom features and visualizes the names of the classes to promote the comprehension at fine-grained level. Although the metaphor is for any object oriented programming language, the current implementation of the prototype supports Java and C++. The prototype can be however easily extended to visualize software systems implemented with different programming languages. It is only needed to implement a tool for extracting the considered metrics from code implemented with those programming languages.

In order to assess the validity of the approach and CodeTrees, we have conducted a preliminary empirical investigation as a case study. We used medium open source software systems. Two of these systems were implemented in Java and one in C++. The results seem encouraging, but due to the preliminary nature of our investigation caution is needed and replications are required to increase our awareness on the achieved findings.

There are several future directions for our work. We plan to conduct users’ studies to evaluate the effectiveness of our proposal in the execution of maintenance or software comprehension tasks. We will use recent guidelines [23] for such a kind of studies. We also plan to improve the tool support in terms of interactions between the maintainers and the virtual natural environment using the idea of Visual Information Seeking Mantra “Overview, zoom and filter, details-on-demand” [24]. We are also going to investigate our metaphor in the visualization of evolving software systems as growing forests. Finally, future work will be devoted to increase the realism of the forest and to assess whether users perform maintenance tasks in a more efficient fashion in case of more realistic forests. The assessment of efficiency will represent another key factor in that investigation.

Acknowledgment

We would like to thank the Manuela Viggiani and Nicola Capece, who developed some of the software modules of the prototype implementing the metaphor presented here.

6. REFERENCES


