Assessing the Comprehension of UML Class Diagrams via Eye Tracking

Shehnaaz Yusuf
Huzefa Kagdi
Jonathan I. Maletic
Motivation

- Traditionally UML usability studies have been performed using methods such as surveys and questionnaires
  - Explicitly collected data after the fact
  - Disparity the reality and what is reported
  - Subject may forget after a lengthy task

- Our study uses an eye-tracker to collect a subject’s activity while performing a given task on UML class diagrams
  - Data is collected implicitly while subject is addressing the task
  - Collects actual eye movements – where the subject was looking and for how long
Eye Tracking Equipment

- Tobii 1750 eye-tracker
  - www.tobii.se
  - Non-intrusive
  - Two cameras built into 17 in flat screen
  - Error rate < 0.5 degrees
  - Sampling rate 50MHZ
  - Records audio and video of session

- ClearView software
  - Gaze plots
  - Heat maps
Name the entity class that is responsible for storing data?
Gaze-Plot

Fixation

Saccades

Name the entity class that is responsible for storing data?
Heatmap (for all subjects)
Our Main Goals

- Assess the effectiveness of different layouts (of UML Class diagrams)
- See if labeling of classes with their stereotype information is helpful
- See if color is useful
Some Basic Questions

- What do people really look?
- Where do people fixate most?
- How do people navigate?
- Is there a difference between experts and novices?
The Study

- Software system used in study
  - Hippodraw – C++ and QT
  - 240 classes (50 KLOC)
- Number of subjects: 12
- Number of tasks: 27
  - 12 which address UML notation
  - 15 which address design problems
- Three different layout strategies [VISSOFT’05]
Orthogonal Layout

Focuses on the minimization of the edge crossings and bending.
Three-Cluster Layout

Positions classes into three clusters (i.e., boundary, control, and entity) based on their design or architectural roles.
Multiple-Cluster Layout

Related classes that are responsible for a specific functionality of a software system are positioned in a single cluster.
## Tasks (Questions)

<table>
<thead>
<tr>
<th>No.</th>
<th>UML Question (1-12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identify the kind of relationship between class <em>ViewBase</em> and class <em>PlotterBase</em>.</td>
</tr>
<tr>
<td>2</td>
<td>Name the classes involved in aggregation.</td>
</tr>
<tr>
<td>3</td>
<td>Name the derived classes of the class <em>PlotterBase</em>.</td>
</tr>
<tr>
<td>4</td>
<td>Name the class with the method name <em>getAverage</em>.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>Design Question (13-27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Name the class that a python wrapper uses to access data in the class <em>NTuple</em>.</td>
</tr>
<tr>
<td>14</td>
<td>Name the class responsible for managing XML serialization?</td>
</tr>
<tr>
<td>15</td>
<td>Name the class that controls the active window of an application.</td>
</tr>
<tr>
<td>16</td>
<td>Name the base class for axis representation hierarchy</td>
</tr>
<tr>
<td>17</td>
<td>Name the class through which a boundary class could access data in the class <em>NTuple</em>.</td>
</tr>
</tbody>
</table>
Subject Accuracy Results

The chart shows the accuracy results for different subjects in terms of Correct Answers. The bars are color-coded to represent Design (green) and UML (orange). The subjects are labeled as K, L, J, A, D, C, H, B, E, G, I, and F. The accuracy values range from 0 to 25.
Subject Classification

Subjects were classified into four different groups based on their accuracy in answering questions:

- UML and design novice
- UML expertise but inexperienced designer
- UML expertise and knowledgeable designer
- UML and design expertise
Classification of Questions

- Based on the distribution of subjects answers (to design questions)
- Ranges [0%, 25%), [25%, 70%), [70%, 80%), and [80%, 100%]

<table>
<thead>
<tr>
<th>Level</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>15, 16, 19, 21</td>
</tr>
<tr>
<td>Intermediate</td>
<td>13, 14, 22, 24</td>
</tr>
<tr>
<td>Difficult</td>
<td>17, 25, 26, 27</td>
</tr>
<tr>
<td>Challenging</td>
<td>18, 20, 23</td>
</tr>
</tbody>
</table>
Analysis & Findings

- **Exploration** of visual space: How they perform searches on the UML class diagram to locate objects required for the task.

- **Examination** of visual objects: How they visualize, in detail, whole or parts of classes and relationships while accomplishing a given task.

- **Navigation**: How they move from one object of interest to the next after their discovery.
Exploration and Examination

- The eye-tracker captures fixations at the granularities of class, attribute, and method textual names in the diagram.
- The first fixations were found only on the end of relationship symbols:
  - (e.g., diamond edge for aggregation) for questions regarding or involving relationships.
- Only saccades were found on the rest of a relationship symbol (i.e., the lines):
  - Line parts of relationship notations are used only for navigation purposes.
Expert vs. Novice

- Expert subjects tended to start exploring the diagrams from the center and moved towards the periphery.

- Other subjects tended to explore the diagrams from top-to-bottom, and left-to-right.
Stereotype Usage

- Experts
  - Examined the textual annotations used for stereotypes
  - Used class colors to facilitate exploration and navigation

- Inexperienced and novice
  - Did not use stereotypes
  - Explored and examined almost all classes in the diagram
Stereotype Usage

Experts divided the visual space into clusters based on the stereotype color

- Used clusters as units of navigation (and not classes)
- Narrowed down search to the cluster potentially containing the answer and examined in detail
- Used the above strategy to answer questions correctly and quickly
Efficient Layouts

- Eye-tracking metrics can be utilized for objective assessment of class diagram layout

- Larger number of fixation
  - Classes and relationships are laid out in a way that leads to inefficient visual exploration, explanation, and navigation
  - Spans the attention of the subject across a number of objects instead of narrowing down to relevant area of interest
Average Number of Fixations

- Relative effort required for tasks are formed from average number of fixations

- The stimuli with average number of fixations in the a classified:
  - [0, 34) - low
  - [34, 42) - intermediate
  - [42, 50) - high
  - [50, 67) - extreme
Classification of Effort based on Average Number of Fixations

<table>
<thead>
<tr>
<th>Stimuli (Questions)</th>
<th>Average Fixations</th>
<th>Effort</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>23.00</td>
<td>Low</td>
<td>Easy</td>
</tr>
<tr>
<td>23</td>
<td>23.56</td>
<td>Low</td>
<td>Challenging</td>
</tr>
<tr>
<td>14</td>
<td>34.33</td>
<td>Medium</td>
<td>Intermediate</td>
</tr>
<tr>
<td>1</td>
<td>36.22</td>
<td>Medium</td>
<td>Easy</td>
</tr>
<tr>
<td>24</td>
<td>42.56</td>
<td>High</td>
<td>Intermediate</td>
</tr>
<tr>
<td>25</td>
<td>42.78</td>
<td>High</td>
<td>Difficult</td>
</tr>
<tr>
<td>13</td>
<td>62.44</td>
<td>Extreme</td>
<td>Intermediate</td>
</tr>
<tr>
<td>27</td>
<td>65.22</td>
<td>Extreme</td>
<td>Difficult</td>
</tr>
</tbody>
</table>
Comparisons on the Level of Questions to the Effort

- The baseline of comparison is that the level and effort should be directly related (*equal-effort*)
  - Easy questions should require low effort, difficult questions should require high effort..etc
- Questions that require more effort than the corresponding baseline level (*more-effort*)
- Those that require less effort than the corresponding baseline level (*less-effort*).
Distribution of Questions based on Level and Effort

<table>
<thead>
<tr>
<th>Layout Types</th>
<th>Equal-Effort</th>
<th>More-Effort</th>
<th>Less-Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthogonal</td>
<td>3</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Three-cluster</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Multiple-cluster</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Equal-effort**
  - Multiple-cluster highest while orthogonal lowest
- **More-effort**
  - Multiple-cluster lowest while orthogonal highest
- **Less-effort**
  - Multiple-Cluster highest in less-effort
Related Work (UML)

- Effective layouts (Sun’05, Andriyevska’05, Eiglsperger’03)
- Influence of stereotypes (Kurniaz’05, Staron’05)
- UML aesthetics (Eichelberger '02, '03, Purchase ’01, ’02)
- Usability studies (Tilley’03, Arisholm et al. ’06, Briand’05)
- Eye tracking on Source code (Bednarik ’06, Uwano ’06)
- Mental workload (Iqbal’05)
- Web page viewing (Beymer’05, Nakamichi’06, Pan’04, Whalen’05)
- Eye tracking research (Jacob ’90, Duchowski ’03, Hyona ’03)
- Eye tracking on UML (Gueheneuc ’06)
Conclusion

- One of the first studies to use eye-tracking equipment
- Proposed the creation of objective assessment metrics of class diagram layout
- Findings showed that experts tend to use such things as stereotype information, coloring, and layout to facilitate more efficient exploration and navigation of class diagrams
- Observation: Close similarity in the notations for generalization and aggregation relationships could cause undue effort to differentiate