Visibility optimization for data visualization: A Survey of Issues and Techniques

Ch Harika, Dr. Supreethi K.P
Student, M.Tech, Assistant Professor
College of Engineering, Jawaharlal Nehru Technological University, Hyderabad, Andhra Pradesh, India
harikachukka24@gmail.com, supreethi.pujari@gmail.com

Abstract
Techniques of visualization use colors to represent categorical differences to the user. Color appearance changes with size of the object being viewed. While using colors, the structures which are more prominent as in position suppress small non uniform groups or inhomogeneous groups. Without maintaining sufficient distance among the colors they are difficult to be discriminated. Class visibility is used to measure the color palette utility and color optimization algorithm based on class visibility makes the differences clearly visible to users.

1. Introduction
Data visualization is a visual representation or graphical representation of data. Graphical representation of data assists the users to recognize patterns, to detect trends, and to identify relationships hidden in data sets. Information that has been abstracted in some semantic form including attributes or variables for the unit of information like temporal data, persistent data can also be created as a study of visual representation. There are few concepts for the selection of visualization to our data. We should understand the data we are trying to visualize, along with its size and cardinality. One must know what we are trying to visualize and what kind of information we want to communicate. The visual that conveys the information in the best and simplest form to the audience must be used. [3]

Data can be represented in different forms of graphs like line graphs, bar graphs, Scatter plots, pie charts etc. Line graphs show the relationship of one variable to another. These are also useful to compare multiple items over the same time period. Bar Graphs are used for comparing the quantities of different categories or groups. Scatter Plot is a good way to visualize relationships in data. Pie Charts are used to compare the parts of whole. This is most effective when there are limited components and when text and percentages are included.[3]

Data visualization concept is mostly useful for representing the categorical differences and these categorical differences can be world climate, vegetation class changes etc.....Selecting categorical colors is challenging problem in data visualization. To choose proper categorical color we use visualization packages as a guidance to choose color. Choosing the categorical data without considering the above aspects can results in misleading of interpretation of data. The problems that generally occur during the selection of categorical colors are: The structures which are having more power and influence over others or more prominent as in position suppress the small inhomogeneous groups. Without maintaining sufficient distances among the colors they are difficult to be discriminated. Color appearance also changes based on spatial variations of data in an image because of visual sensitivity of humans. Color design also forms a crucial part in visual aesthetics. [9]

The problems that occur during selection of categorical colors can be addressed by class visibility concept. By using this class visibility we measure the utility of color palette to make the categorical colors clear to the users. Visual saliency will make the will make the more prominent structures and small inhomogeneous groups clearly visible to users considering spatial variations. The visual saliency estimation relies on spatial averages, color averages. Apart from using the visual saliency, the colors do not correctly label the categorical groups like narrow areas with high visual sharpness. Point saliency is also used to measure the intensity of single class in pixels. The class visibility and point saliency both together will make the categorical groups differentiate.

Visibility based optimization is used to improve the color of categorical data visualization. Visual based algorithm will increase the intensity of small groups and reduce the weight of large groups. Optimization procedures will maintain distinguishable distances between the colors. [1]

A Typical method of visualizing data sets involves mapping data attributes to visual features like shape, size, spatial location, and orientation. Color is the important frequently used feature. The general rule of thumb for computer generated images is not more than five to seven colors. [6], previous works has addressed to increase
the total number of colors available by taking three aspects under consideration during selecting: 1. Color
distance, 2.Linear separation, and 3.Color category. Also the results have showed that the selected group of
colors will provide good differentiation between the data elements during data visualization [6].

2. Methods of Data Visualization
   Different methods are available for visualization of data based on the type of data. Data can be: [2]
   - Univariate
   - Bivariate
   - Multivariate

1.1 Univariate
   It is a measurement of single quantitative variable. It is Characterised distribution which is represented
   using following methods.

   **Histograms:**

   ![Histogram](image1)

   **Pie charts**

   ![Pie chart](image2)

1.2 Bivariate
   This constitutes a paired sample of two quantitative variables. Variables are related and are represented
   using the following methods.

   **Scatter Plots**

   ![Scatter plot](image3)

   **Line Graphs**

   ![Line graph](image4)

1.3 Multivariate
   This is a multi dimensional representation of data and is represented using following methods
3. Effective color palettes

3.1 Color Selection Principle
The colors are selected by using the principal of order and separation. Order will arrange the colors in sequential order and separation wants all the colors used to make the differences must be different. There are many approaches that were proposed to realize order and separation for the colors that are isoluminant [2][4]. In some situations many users can select their favourite color but they do not know the contrast of categorical groups. So, this made to improve the color mapping for users.

3.2 Size/Frequency Dependency
Visual sensitivity of the human visual system differs in spatial frequencies of data items in an image.

3.3 Feature driven attention model
This is a 2 stage process in the feature integration theory. First stage: Preattentive primitives like color, lightness are detected and are separately encoded into feature maps. These stages are performed rapidly in parallel. Second stage: This is a slow serial conjunctive search. This will integrate or combine the future maps into single topographical saliency map. [1]

3.4 Drawbacks
The colors appearances changes with size of the object being viewed. The interaction between the size and the color appearance is a critical design factor. By considering 5 colors of different hues of the same luminance. By using these 5 colors in both line graph and bar graph, the difference can be easily distinguished as shown in figure 1. [10], the colors used in first graph are clearly visible than the colors used in second graph.
In Figure 2 the shaded colors are easily distinguish in the table but difficult to distinguish in the small color picker.[10]

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2,997.16</td>
<td>-8%</td>
<td>4,643.98</td>
</tr>
<tr>
<td>7</td>
<td>4,855.08</td>
<td>34%</td>
<td>5,059.35</td>
</tr>
<tr>
<td>8</td>
<td>9,105.70</td>
<td>30%</td>
<td>3,690.86</td>
</tr>
<tr>
<td>9</td>
<td>8,243.00</td>
<td>0%</td>
<td>7,675.18</td>
</tr>
<tr>
<td>10</td>
<td>6,557.27</td>
<td>-4%</td>
<td>7,230.73</td>
</tr>
<tr>
<td>11</td>
<td>9,521.20</td>
<td>2%</td>
<td>9,992.54</td>
</tr>
<tr>
<td>12</td>
<td>1,962.02</td>
<td>5%</td>
<td>9,952.49</td>
</tr>
<tr>
<td>13</td>
<td>3,867.51</td>
<td>50%</td>
<td>681.20</td>
</tr>
<tr>
<td>14</td>
<td>8,313.33</td>
<td>25%</td>
<td>2,298.35</td>
</tr>
<tr>
<td>15</td>
<td>777.77</td>
<td>-8%</td>
<td>6,457.72</td>
</tr>
<tr>
<td>26</td>
<td>4,493.30</td>
<td>7%</td>
<td>6,899.20</td>
</tr>
<tr>
<td>27</td>
<td>5,302.02</td>
<td>33%</td>
<td>2,961.35</td>
</tr>
<tr>
<td>28</td>
<td>351.36</td>
<td>-3%</td>
<td>3,645.93</td>
</tr>
<tr>
<td>29</td>
<td>247.02</td>
<td>10%</td>
<td>3,241.45</td>
</tr>
<tr>
<td>30</td>
<td>5,057.84</td>
<td>47%</td>
<td>5,388.82</td>
</tr>
<tr>
<td>31</td>
<td>5,203.66</td>
<td>44%</td>
<td>2,297.25</td>
</tr>
<tr>
<td>32</td>
<td>200.51</td>
<td>30%</td>
<td>2,048.94</td>
</tr>
<tr>
<td>33</td>
<td>9,040.79</td>
<td>7%</td>
<td>2,508.59</td>
</tr>
</tbody>
</table>

**Fig 2**

In Figure 3 unlike the above 2 diagrams a similar blending process occurs in human visual system which is called as effect spreading While seeing that colors the human eye blends at small areas of the colors that are close together even though the same colors are used in both bars.[10]

**Fig 3**

In Figure 4, the structures which are more prominent suppress small inhomogeneous groups.[1]
4. Research Goals
Our optimization tries to make the categorical differences clearly visible to the users by increasing the intensity of small groups and by reducing the weights of large groups.\[1\]

That means optimization tries to increase the contrast of small regions and reduce it for large regions which will allow to produce image with balanced colors, according to the associate size of the categorical groups. [1]

5. Class Visibility and point saliency
This is the measure of how the color and spatial distribution of each class affect the intensity of human visual system. Class visibility is estimated by weighting its point saliency with its structural weights. The structural weights are calculated by the contributions of neighbouring points (Derived rather than measure). [1][9][10]

Unlike the conventional filtering we only collect the contributions of points whose labels are the same as the centre points. By preventing or excluding the contributions of points of different labels, we can avoid losing categorical differences.

Point saliency will exactly weight the contribution of categorical color at a single data point. The point saliency is measured as the color difference from the average color of neighbouring data points.

\[
S_p = \Delta \epsilon \left( C_p \frac{1}{|N_p|} \sum_{q \in N_p} C_q \right)
\]

Color difference, $\Delta \epsilon$, is measured by the CIE76 metric, which is the Euclidean distance in CIE Lab.\[1\]

Our class visibility defined for each data point, is estimated by weighting its point saliency with its structural weights.

\[
V_p = w_p S_p,
\]

where $w_p$ is a structural weight of $p$ that measures the contribution of $c_p$ to the visibility of $c$’ in $\Omega_p$. 

The structural weights, effectively captures the difficulty of a user to observe a specific class at narrow local regions.

6. Color Optimization

6.1 Mean class visibility
We need a representative measure of each class. The mean class visibility of each class is computed by the average of the data points that belong to its category.

$$V^m = \frac{1}{|c^m|} \sum_{p \in c^m} V_p,$$

where $|c^m|$ is the number of its data points.

6.2 Cost Function
Cost function is defined for each data point is to minimize differences among the class visibility levels in its local surroundings.

$$E(X) = \frac{1}{M} \sum_{m=1}^{M} (V^m(X) - T)^2$$

Where,
- $M$ = Number of classes
- $V^m$ = Mean class visibility of class $M$
- $X$ = Color Vector
- $T$ = Target Vector

6.3 Optimization
The final resulting cost functional with the constraints to optimize is:

$$F(X) = E(X) + \lambda [B(X) + R(X)]$$

Where,
- $B(X)$ = Bounding constraint
- $R(X)$ = Color discrimination constraints (colors should maintain sufficient perceptual distances).

7. Conclusion
The color optimization algorithm based on visual sensitivity of human makes the categorical differences clearly visible in graphical representation of data. Also to make more prominent structures do not suppress small inhomogeneous group.

8. References


