Media Visualization of Book Cover Images: Exploring Differences among Bestsellers in Different Countries

Wooseob Jeong
School of Information Studies
University of Wisconsin - Milwaukee
P.O.Box 413 Milwaukee, WI, USA
wij8612@uwm.edu

Hyejung Han
School of Information Studies
University of Wisconsin - Milwaukee
P.O.Box 413 Milwaukee, WI, USA
hanh@uwm.edu

ABSTRACT
Interest in the role of book cover images in readers’ reading experience and book marketing has been long-standing. This study attempts to compare book covers from different countries with a media visualization tool called ImagePlot. The top 100 bestselling books from 13 Amazon.com’s international sites were identified and their cover images were downloaded. Using ImagePlot, median values of brightness, hue and saturation for each image in the data set were extracted and analyzed. Along with one-way ANOVA tests by SPSS, ImagePlot outputs show differences in these graphical properties of bestsellers’ cover images in different countries. From the outputs, with all the book cover images displayed on a single canvas (screen), hidden patterns emerged and findings were clearly confirmed. This study makes a contribution by providing connection between research interests in book cover images and media visualization techniques for further research.

Keywords
Media visualization; visualization of image collection; book cover images; cultural analytics; digital humanities

INTRODUCTION
Contrary to the real meaning of the English idiom, “Don’t judge a book by its cover,” research interest in book covers has been steady ranging from the personal book selection process to the marketing of trade books. For example, there is an edited scholarly monograph titled “Judging a Book by Its Cover” (Matthews & Moody, 2007) discussing book covers’ important roles. In fact, the importance of cover images was clearly identified from the late 19th century, although 1930’s Penguin Books’ paper revolution is generally considered as the beginning of the interest in book covers (McCleery, 2007). For example, from the 19th century, book bindings were commonly decorated on their covers or dust jackets to reflect the content of the book, which is a modern development (Pearson, 2013) and around the same time, the growing interests in children’s book publishing started to emphasize the importance of book covers to attract the readers’ eye and to promote the books for sales (Immel, 2013). Despite abundant research activities on book covers, however, most studies are based on individual book cover images, often with different cover images of a single title over time and in different countries. A new method called media visualization can help to overcome this limitation, allowing researchers to explore various properties of images in a collection or to compare among multiple image collections.

PURPOSE OF STUDY
The purpose of this study is, utilizing a media visualization method by ImagePlot, to find if there is any difference in brightness, hue and saturation of cover images of bestseller books from different countries and to identify any patterns among the differences and if so, what the characteristics of the differences may be.

METHODOLOGY
Data Collection
In mid-April, 2014, from Amazon.com’s 13 international sites, we collected the cover images of top 100 bestsellers. The 13 international sites include: Australia, Brazil, Canada, China, France, Germany, India, Italy, Japan, Mexico, Spain, United Kingdom and United States. A few books did not display their cover images. Table 1 shows the number of images in each country’s data set with each country’s country code. Some images have white borders around actual cover images or “Look Inside!” icons, but due to time constraints, those images were not cropped, but used in the data process as they were.
Data Processing
The collected images were saved in a designated folder for each country. By “ImageMeasure” macro, median values of hue, brightness and saturation for each image were extracted and saved in a corresponding data file for each country. For the overall visualization and analysis, all the images regardless of their country origins, were collected in a separate folder. Their hue, brightness and saturation values were measured again and saved into another data file. In this file, the Internet country code was assigned for each data entry.

Visualization Tool - ImagePlot
ImagePlot was developed by Manovich and his lab colleagues with various macros. In it, all the values measured are converted into 0-255 scale (ImagePlot Documentation). For brightness measurements, color images are internally converted greyscale using this formula: gray=(red+green+blue)/3. If an image is pure black, Brightness_Median will be 0; if an image is pure white, it will be 255. Low Saturation_Median value indicates that image colors are mostly desaturated; a high value indicates that most colors are close to being pure (very saturated). The 0-255 scale may lead to some confusion particularly for hue values since hue values are normally represented as degree of a circle, 0 to 360, from yellow to red, with green, blue and purple in the middle. Therefore, when the hue values are represented in a linear form, such as X-axis or Y-axis, the interpretation requires extra caution.

Data Analysis
For the overall image collection, 3 sets of visualization outputs were rendered by ImagePlot: brightness by country, hue by country and saturation by country. For the differences among countries, one-way ANOVA analysis was conducted with means and standard deviations (stdev) of the measured median values of brightness, hue and saturation. Bar graphs of the result were generated by the SPSS software. For combined analysis, 3 sets of bivariate graphs were generated for each country’s collection by ImagePlot: brightness vs. hue, brightness vs. saturation and saturation vs. hue. The output bivariate graphs were compared among countries for further analysis.

RESULTS
Differences in brightness, hue and saturation by countries (ANOVA)
A one-way ANOVA test was conducted by SPSS with the median values and standard deviation values of each image for its brightness, hue and saturation. The results showed that both the median values and the standard deviations are significantly different statistically (p<.05) among countries for all three of the properties.

Differences in brightness, hue and saturation by countries (Media Visualization)
The outputs from ImagePlot provides more in-depth, but still easy to identify, representation of the book cover images among countries. The three graphs in Figure 4 show the different brightness, hue and saturation of each book cover image from the data by country respectively. All the books in the data set were represented on the individual graph, with actual images of the covers with the size of 50 by 50 pixels. The X-axis shows each country represented by its Internet country domain name, and the Y-axis shows brightness, hue and saturation on a scale of 0-255. For brightness and saturation, the lower the Y value, the darker or the less saturated each image is. For hue, Y values represent, from lower to higher, yellow, green, blue, purple, red colors.

In this method, since all the actual book cover images are displayed on a graph, it is easy to identify patterns throughout the whole data. For example, Chinese and Japanese books do not have many dark images on their covers and French, German, Indian and British books have more red/purple colors on their covers. Most book covers do not have highly saturated colors, but Japanese books distinctively do not have highly saturated colors on their covers.

\begin{table}
\centering
\begin{tabular}{|c|c|c|}
\hline
Country & number of Images & Country & number of Images \\
\hline
Australia (AU) & 100 & Italy (IT) & 100 \\
Brazil (BR) & 100 & Japan (JP) & 94 \\
Canada (CA) & 100 & Mexico (MX) & 100 \\
China (CN) & 100 & Spain (ES) & 99 \\
France (FR) & 100 & United Kingdom (UK) & 100 \\
Germany (DE) & 100 & United States (US) & 99 \\
India (IN) & 99 & & \\
\hline
\end{tabular}
\caption{Number of Images by Country}
\end{table}
Figure 1. Brightness, Hue and Saturation by Country

Bivariate Analysis (Media Visualization)

Brightness vs. Hue
Figure 2 shows four ImagePlot outputs for brightness and hue in the image data sets from four countries, China, Japan, United Kingdom and United States. X-axis is for brightness, the higher the value, the brighter the image, and Y-axis for hue, from lower value to higher, representing yellow, green, blue, purple and red. These outputs show that Chinese and Japanese book covers are mostly brightest and less colorful, while British and American book covers are not only various in color but also diverse in brightness.

Brightness vs. Saturation
Figure 3 shows four ImagePlot outputs for brightness and saturation in the image data sets from four countries, China, Japan, France and India. X-axis is for brightness, the higher the value, the brighter the image, and Y-axis for saturation, the higher the value, the more saturated. In terms of brightness, French and Indian books have more dark colors on their covers than Chinese and Japanese books. They also have more saturated colors on their covers than Chinese and Japanese books, meaning Chinese and Japanese book cover images have more white, black or gray elements in their colors.

Figure 2. Brightness (X-axis) vs. Hue (Y-axis)

Figure 3. Brightness (X-axis) vs. Saturation (Y-axis)
subtle distinctions are explored without Chinese or Japanese books.

Figure 4 shows four ImagePlot outputs for hue and saturation in the image data sets from four countries, Germany, Italy, Mexico and Spain. X-axis is for saturation and Y-axis for hue. From these graphs, it is reasonable to conclude that German and Italian books have more red/purple colors on their covers than Mexican and Spanish books (more blue colors on Mexican and Spanish book covers as well), while Mexican and Spanish books have more saturated colors relatively than German and Italian books.

DISCUSSION

Differences among countries and cultures
In the previous section, ANOVA tests and the media visualizations by ImagePlot showed that there are clear differences among the cover images of bestselling books in different countries, in terms of their brightness, hue and saturation values. The differences are clear enough to allow us to draw a cautious but reasonable generalization regarding cultural difference in book cover images. For example, between East Asian culture and Anglo-American culture, there are distinctive differences in brightness, hue and saturation. East Asian book covers tend to have brighter colors, fewer red/purple colors, and fewer saturated colors than Anglo-American books covers. These differences may reflect different consumer preferences in book cover images, which gives a good insight for global book translation business.

Potentials of media visualization in information science
Information science has a long and strong tradition of information visualization such as bibliometrics and social network analysis. Media visualization can enrich this tradition by enabling more cultural and historical artifacts data to be analyzed with flexible graphical analysis technique. Since media visualization allows direct display of images of artifacts or still shots of video, with all the data displayed on a big canvas (screen), it can reveal unknown patterns, which may not be visible if we only use a small number of cases (Manovich, Douglass & Zepel, 2012).

Study limitation and future research direction
It would be more desirable if a much larger data set is used. Including more countries would be better as well, particularly if cultural differences are pursued. Even if changing cover images when translated into other languages or countries is common, in many cases the original cover images are used in multiple countries. It would be interesting to see how much the result may differ, if only unique cover images are included. However, if time is considered as a variable, it is also necessary to consider the fact that even in the same country a book may change its cover images numerous times for marketing purposes.

Expansion of this study into other image collections is natural. For example, jacket images of music CD or DVD collections can be easily analyzed by the same technique used in this study. The analysis can be conducted in multiple ways including by time or location.

CONCLUSION

It is believed that this study makes a contribution by providing connection between research interests in book cover images and media visualization technique for further investigation. In practical application aspects, it also contributes in the book business, particularly in the area of translation and marketing of books in different countries with different cover images.

REFERENCES