Colors of the street: color as an image visualization parameter of Twitter pictures from Brazil's 2013 protests

Johanna I. Honorato
johonorato@labic.net

Lucas O. Cypriano
lucascypriano@gmail.com

Fábio Goveia
fabiogoveia@labic.net

Lia Carreira
liacarreira@labic.net

Labic, Universidade Federal do Espírito Santo
514, Fernando Ferrari Ave. - Vitória, ES – Brazil – CEP: 29.075-910
+55 27 4009-2752

ABSTRACT
This paper aims to discuss color as a methodological tool in the analysis of large quantities of images. For this purpose, this paper presents a series of researches done by two data analysis labs, Software Studies Initiative (EUA) and Labic, the Laboratory of Image and Cyberculture Studies (Brazil), in order to illustrate its different uses. Moreover, this paper shows Labic's recent research on color as a parameter for the analysis of 85.585 images linked to twitter hashtag #vemprarua, an important hashtag related to Brazil's 2013 protests. Thus, this paper highlights the importance of colors as parameters, while identifying issues and contributions to contemporary data science.

Categories and Subject Descriptors
I.4.8 **[Image Processing And Computer Vision]***:* Scene Analysis *– color.*

General Terms
Measurement, Documentation, Design, Standardization.

Keywords
Big Data, Colors, Data visualization, Image, #Vemprarua, Image analysis.

1. INTRODUCTION
The production, dissemination and storage of digital images have achieved large scales with rapid technological advances and accessibility in contemporary society. Image production, with its multiplying variety of tools and available apps for online sharing, has boosted this ever changing scenario, being, therefore, an important and complex contemporary context to be studied and better comprehended.

Differently from contemporary semantic studies (that already is a well developed research field, with its well established tools and softwares), the analysis of large amounts of images is still underexplored, considering that there are fewer tools and researches presently available regarding image datamining, visualization and analysis. Image processing and storing requires great memory capacity and powerful devices, as well as specialized professionals. Although in recent years these processes have become more accessible to all sorts of researchers (with its vast developments and lower prices), extraction and analysis of large amounts of images remains a challenge due to its peculiarities.

In this research scenario, images are analyzed though different data parameters, such as its sharing frequency, time and/or size, in order to create all sorts of visualizations. However, this paper focuses on researches that use different types of color information (such as hue, brightness and saturation) as a parameter for analysis and visualization of image data[1]. Our goal here is to study its importance in revealing a variety of patterns and dissonances that can help us better understand the context and modes of image production today.

Thus, our paper focuses on data collected from the 2013 Brazilian protests #vemprarua hashtag on Twitter, retrieved from the 15th of june to the 15th of July of the same year. The 2013 protests became a large movement that gained the support and participation of millions of people in the whole world. With this large engagement, social media websites gained great relevance, enabling protesters to rapidly share pictures and ideas, and promote a variety of debates and events. It also enabled people at home to become part of this social movement, sharing information and, thus, helping to promote the event and spread the news. Due to its importance to Brazil's social and political context, this paper also aims to better understand its contexts and repercussions though image analysis using color parameters.

2. INTERNACIONAL STUDIES USING COLOR AS A METHOD FOR IMAGE ANALYSIS IN BIG DATA RESEARCH

2.1 Color Analysis in Visual Arts
Visual art, such as paintings, can be one of many spheres in which patterns can be revealed though color analysis. For example, painters make use of a variety of colors to produce their works of art and establish themselves within a specific artistic style. Using color as parameters when creating visualizations of these art works, we can perceive and analyze certain differences between different artists and their works, enabling comparative analysis or even analyze what can be called "stylistic development" of a particular painter.

On this matter, Software Studies Initiative published in June 2011 a research analyzing two visual art collections, one by Piet Mondrian and the other by Mark Rothko. The research was based on their images' visual elements (such as hue, brightness and saturation), thus revealing patterns not only between the works themselves, but also between the artists. The purpose of that particular study was to compare a certain number of Mondrian's paintings to Rothko's produced in similar periods of time in their

Conference'10, Month 1–2, 2010, City, State, Country.

[1] This paper, due to its size limits and piratical purposes, does not aim to present an analysis of the context of visual studies and visual perception, although Labic recognizes its importance to the field.

careers. Through this comparison, the research identified their initial predominant artistic style as being related to the styles of their predecessors. But it also found that, as the years went by, Mondrian and Rothko began to differ their color pallet, which was interpreted by the research as the artists' concerns in developing their own style, therefore, diverging from figurativism[2].

Another image analysis was done by Software Studies Initiative using color in relation to time parameters. That particular study was based on Van Gogh's experience in Paris compared to his time in Arles, and it showed that the set of images from the later contrasted with the set from the former, due to its higher saturation and brightness - a result of the painter's new color experimentations. Thus, the visualization proposed by Software Studies suggests that Van Gogh's paintings were influenced by the spatial changes in his life, as he moved from one city to the other.

2.2 Phototrails' Color Analysis

In July 2013, Nadav Hochman, Lev Manovich and Jay Chow - researchers from Software Studies Initiative -, developed a project called Phototrails. Their goal was to explore, in a planetary scale, visual and dynamic patterns and structures of user-generated contents of Instagram. The study showed, thorough visualizations created using images from that photo sharing online network, how temporal changes and visual features of different locations can reveal their social, cultural and political characteristics, as well as people's habits around the world. In one of their analysis, the researchers chose, among millions of images captured from Instagram, several random samples of various cities, each containing 50,000 images. From that chosen dataset, it was extracted basic visual information (such as average color, brightness, saturation, number of edges, contrast, etc.) to create different visualizations and, thus, highlight each city visual identity in a specific period of time[3].

2.3 Flickr Flow's Color Analysis

Flickr Flow is a project from 2009, developed by data visualization researchers Fernanda Viégas and Martin Watternberg, which serves as an example of image visualization by color using contemporary photographs to retrieve its data. The study started using collections of photographs of Boston Common found and extracted from the photo social network Flickr. With their available data, the researchers divided all photos by month and calculated their colors' relative proportions. The projects following step was to, then, plot a "wheel" shaped dataviz using both color and time as its parameters[4].

Thus, as a result of the visualization created, differences between the seasons of that particular year can be identified through its color variation pattern. At the bottom of this visualization, it's possible to identify a great amount of grays, whites and lighter colors, which represent winter. One can then observe, clockwise, the increase of more vivid colors (variations of pink, purple, green and yellow), thus representing spring. Following this pattern, one can also observe the other seasons, with fall being indicated by the yellows and oranges and summer by large amounts of bright colors and a very few of white tones.

[2] Images and more infos on the study are available at http://lab.softwarestudies.com/2011/06/mondrian-vs-rothko-footprints-and.html

[3] The research results and images are available at http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698

[4] The research results and images are available at http://hint.fm/projects/flickr/

3. #VEMPRARUA's COLOR ANALYSIS

3.1 2013 Brazilian Protests and its hashtag #vemprarua

The previous studies conducted by Software Studies Initiatives has shown the vast possibilities regarding image analysis using colors as parameters, bringing great contributions to data science. Taking in consideration the contribution that they have also made to the analysis of social and cultural behavior and patterns through image visualization, Labic has been developing, using other tools and visualizations, a research in which we can better understand the complexity and variety of political and social issues implicated in the emergence of June's 2013 protests.

The objective of this study, named "Visagem", is then to analyze Twitter hashtag #vemprarua (which can be translated as "come to the streets"), an iconic expression of the Brazilian protests and, thus, the most used hashtag to refer to this particular social movements within social media websites. The 2013 protests in general were against government corruption, poor government financial administration associated with 2014 World Cup, and also for better quality of transport, security, education throughout the country. It is then an important social and political movement that deserves especial attention and research.

3.2 #Vemprarua's Datamining Process

Our datamining method was based on the retrieval of data from the popular social networking service Twitter through a software called yourTwapperKeeper (a.k.a YTK), which uses Twitter API to gain access and extract the necessary data. With this method, all tweets that had the matching hashtag *#vemprarua* were collected, creating a csv file with all the available information (such as who tweeted, date of publication, number of retweets, etc.).

With this csv file, Labic used a Java based script called *Crawler*, developed by our lab and whose function is to separate tweets that contain links from those that don't. After this process, the script access each tweeted link and captures the images that obeys the parameters set previously by our researchers, such as a minimum size of 15 kB or 200 x 200 px, and the extension files PNG, JPG, JPEG, TIF or TIFF.

Between June 15 and July 15 of 2013 (a critical period in that year's political and social protests), we extracted 85,595 images, originated from a total of 404,006 tweets. These images, despite only being retrieved from Twitter, came originally from a variety of websites and apps, such as online news websites, blogs, and other social networking websites, that was then shared by several social media profiles.

3.3 #Vemprarua's Color Parameters

In order to analyze this large amount of images, HSB color scale was used in this research, in which the color of each pixel in a image is composed by three numeric data: hue, brightness and saturation. Basically, hue values goes from 0 to 255 (equivalent to 0º to 360º degrees), thus forming a color circle. Brightness is then determined by values ranging from 0 to 100, in which zero means no light (black) and 100 means maximum presence of light (white). Finally, saturation follows the numerical variation of brightness, also ranging from 0 to 100, however, being 0 an absence of tone (presence of grays) and 100 being fully saturated colors (no grays). This chosen color scale basically helps identify groups of images with close measurements, and also allows image organization using these same parameters.

With this issue settled, the plug-in "Measure" (a plug-in of the software ImageJ) was used in order to read the values of each pixel and then calculate its hue, brightness and saturation. Thus, it

was through these three color parameters that this study was able to develop different visualizations and analysis of large amounts of images from the #vemprarua movement, enabling the researchers to identify certain patterns and characteristics.

3.4 Analysis of #vemprarua 's Image Visualizations

With the images captured through YTK and with the visual information gathered by the Measure plugin, it was then possible to plot different visualizations in which these large volumes of data can be compared. In order to make these plots, we used a software called ImagerPlot, which was developed by Software Studies Initiatives, housed within the UCSD Division of the California Institute for Telecommunication and Information Technology.

3.4.1 Brightness x Saturation

With these plots, which visually highlights sets of images separated by its color parameters, an analysis was made possible. In the visualization below (Figure 1), #vemprarua's images are distributed throughout three major groups: a whiter set, mostly found in the upper left quadrant; a darker set, found at the base of the this dataviz; and a more colorful set, on the right upper quadrant.

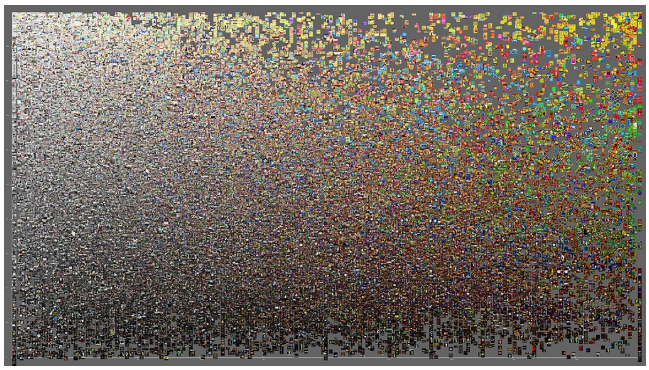

Figure 1 - 85.595 images sorted by X-axis (saturation median) and Y-axis (brightness median)[5]

In the first group with predominantly white images, a greater presence of posters, prints of documents and newspapers covers that have been shared throughout the months in which the protests occurred can be noticed. The distribution and circulation of information of this type was predominant through June 15 to July 15, where, among other contents, it is possible to find: posters aiming to motivate people's participation in the protests, as well as to sharing of more information on the objectives and schedules of the events; documents and newspaper covers that contained information from mass media; and others.

The second group consists in images taken during the protests. They are images of a grayish tone, due to the predominance of the streets' asphalt color, visible during the daytime as well as at night (however, with a darker tone), thus beeing one of the most striking features of these events.

The third group is what aggregates posters and advertisements attached to the contents shared with the #vemprarua hashtag. The posters in this group moves away from the previous black and white pattern, towards more vivid colors: mostly blue, green and yellow, thus largely associated to Brazil's national flag.

3.4.2 Hue x Brigtness and Hue x Saturation

The next visualizations (Figure 2 and 3) arranged the images in #vemprarua's dataset according to its color bands when the parameters were modified to "Hue" (X axis) and "saturation" (Y axis). Thus, groups of similar images are clearly marked and the appearance frequency of certain types of images throughout the collection are better understood. The tracks that stands out are red, orange, yellow, green, blue, and the combination of purple and pink.

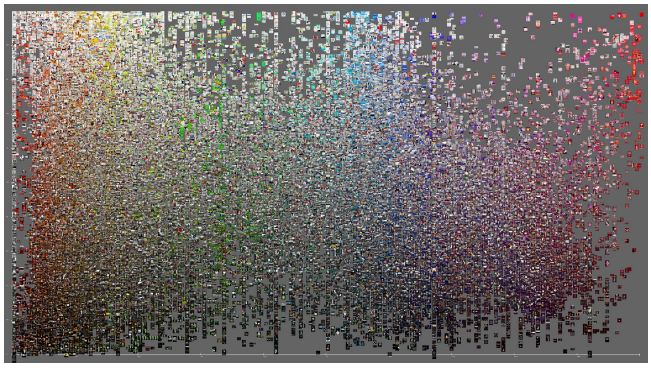

Figure 2 - 85.595 images sorted by X-axis (hue median) and Y-axis (brightness median)[6]

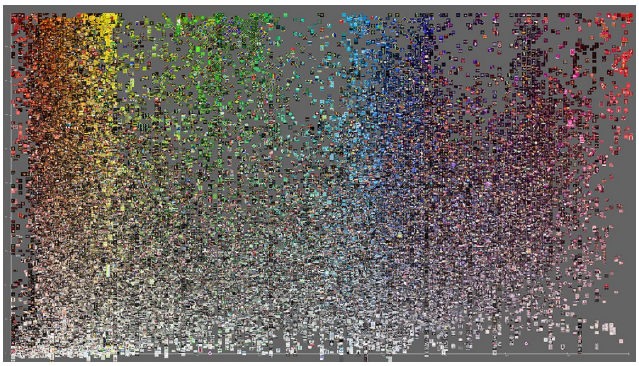

Figure 3 - 85.595 images sorted by X-axis (hue median) and Y-axis (saturation median)[7]

In these visualizations, the first color range has a larger number of images if compared with the rest of the dataset. The predominant images that appears in this group are photos taken at the time of the protests, even if they were shared later on by the users. With a closer look to the predominant orange tone area, images that are characterized by the street's yellow-orange lighting can be observed, as well as photos of confrontation between police force and protesters, which often involved fires being set, explosions and rubber bullets fired by police and captured by the lenses of the vigilant photographers and protesters.

The green color range is basically formed by the reproduction of the national flag of Brazil and also compose by its re-appropriations: these images varies in size, color and type, and occasionally inserted into green colored posters. On some of these posters, the white band that bears the inscription "Order and Progress" (Ordem e Progresso) in Brazil's flag was replaced by,

[5] It's important to notice that the visualizations here presented are created to be visualized in a larger digital visual devices that enables zooming features and user interaction. For a better visualization experience, this image is available in high definition at http://zoom.it/G8jg

[6] Available in high definition at http://zoom.it/QCYi

[7] Available in high definition at http://zoom.it/tAaW

"In Progress" (Em Progresso), meaning that the country was in a state of change led by the people.

The blue color range is also mostly composed by images of flags of Brazil, focusing on its inner circle. This group also has a lot of photos from Instagram, due to one of its available filters. The last color range covering the pink and purple tones are pictures of the protests that were intentionally faded (with the use of filters, for example) and posters intending to represent a more feminine approach.

3.4.3 Color Visualization by Hue with Static Brightness and Saturation

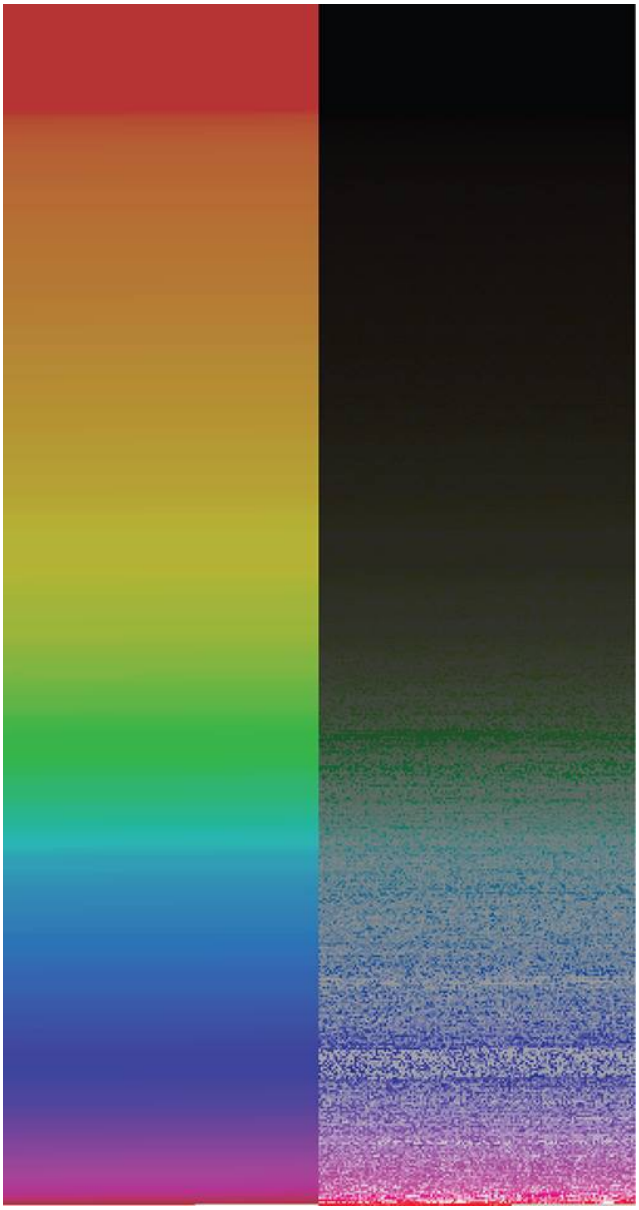

Figure 4a and 4b - Visualization using the median values of hue, saturation and brightness

Instead of placing an image in a specific position determined by its color parameters, we proposed in Figure 4 two different kinds of visualizations. Therefore, for these visualizations we used two types of sets of color parameters. For the first dataviz, we used the hue median, saturation median and brightness median, as to compose the color of the squares representing each image. For the second dataviz, we used just the hue median of each image; the saturation and brightness were established through a standard value. So in Figure 4a we have a visualization of all color parameters of each images, and in Figure 4b it's possible to see more clearly just the hue value. These visualization were created using a script developed in our lab through Processing, in order to visualize the colors medians previously calculated by ImageJ. In these dataviz, each image is thus represented by a square of 2 by 2 pixels, in which the top represents the images with hue median 0 and the bottom, images with hue median 255. Thus, as a result, these recent visualization developed at Labic highlights the large color variations of the Brazilians 2013 social political protests, showing its characteristic visual aspect.

4. CONCLUSION

Throughout this paper, we aimed to understand the importance of the usage of color as parameters for visualizing large amounts of images. Both in the artistic field and in the studies of social movements, color value can reveal more than numeric information: it can also highlight their characteristic visual aspects or styles, as well as point out patterns and singularities of their datasets. As shown in this paper, they can be viewed singularly or in comparison to other images sets. In both cases, color parameters presents themselves as a relevant and simple method for big data analysis.

However, visualizations made with softwares such as ImageJ have certain limitations that can hinder a deeper analysis of the datasets. Using this kind of visualization tools, image plots can only be made using only two coordinates (X and Y). Thus, when a image has the same coordinates as another, an overlapping occurs and you lose, therefore, visual information. Considering this problem, we perceive a need to create a tool capable of adding a Z axis allowing a 3D environment, where image information would not be lost and user interaction is enhanced.

On the other hand, when using Processing, each image is represented by its corresponding pixel, and thus no overlap occurs. In the dataviz presented in 3.4.3 (Figure 4a), each color range in the dataset is clearly represented, confirming the theory that previously observed through the analysis of the first visualizations created with ImageJ (Figure1, 2 and 3), that mostly orange toned pictures were shared during the protests. This fact emphasizes the frequent need of different visualizations of the same dataset for comparison in order to identify or confirm certain characteristics.

The second dataviz in topic 3.4.3 (Figure 4b) follows the same principles of the first, in which the image is represented by its pixels' color, brightness and saturation values(HSB), showing that despite the predominant color being orange, the protests' general tone is dark, which again refers to specific characteristics of the June protests: being an predominantly evening event.

Thus, this paper acknowledges that visual characteristics (such as hue, brightness and saturation), when used as a parameter to organize large amounts of images, can reveal artistic patterns and also social, cultural and behavioral patterns. Therefore, image visualizations using color parameters can present more then numerical values, also pointing to various perspectives of an determined event of practice.

5. ACKNOWLEDGMENTS

Our thanks to the Federal University of Espírito Santo (UFES), the National Council for Scientific and Technological

Development (CNPq) and the Espírito Santo Research Support Fundation (FAPES) for the this research financial support. This research is part of the "Visagem" project of the Laboratory of Image and Cyberculture Studies (Labic).

6. REFERENCES

[1] Hochman, N., Manovich, L., Chow, J. 2014. *Phototrails*. Available at: < http://phototrails.net/>.

[2] Manovich, L. 2012. *Data Visualization and Computational Art History*. Available at: <http://lab.softwarestudies.com/2012/04/data-visualization-and-computational.html>

[3] MANOVICH, Lev. 2011. Style Space: *How to compare image sets and follow their evolution*. Available at: <http://lab.softwarestudies.com/2011/08/style-space-how-to-compare-image-sets.html>

[4] Manovich, L., Hochman, N. 2013. *Zooming into an Instagram City: Reading the local through social media.* First Monday, v.18, n7, 2013. Chicago. Available at <http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698>