

# The color of the street: color as images visualization parameters of twitter pictures from Brazilians Manifestations of 2013

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

Using big data analysis, this paper aims to discuss color parameter as a methodological tool for the analysis of big quantity of images, presenting a series of research done by labs such as the Software Studies Initiative. Moreover, this paper shows recent Brazilian researches by the Laboratory of Studies on Image and Cyberculture (Labic), using color methodology for the analysis of 85.585 images linked to twitter hashtag #vemprarua, related to the 2013 Brazilian protests. Thereby, this work aims to highlight the importance of colors as parameters, regarding researches of this magnitude, in order to identify issues and contributions to studies related to contemporary data modes of productions and sharing.

## 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, hashtag.

## 1. INTRODUCTION

With the rapid technological advances and accessibility in contemporary society, production, dissemination and storage of digital images have achieved large scales. Large 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 has a well developed research field, with well established tools and softwares-, image analysis within big data is still under explored, if we consider the few tools and researches presently available

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

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

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

regarding image datamining, visualization and analysis. Image processing and storing can require 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 and analysts, with its vast developments and lower prices, extracting and analysis of large amounts of images, due to its peculiarities, remains a challenge.

Within this scenario, images can be analyzed through different data parameters when creating all sorts of visualizations, such as its sharing frequency, time, size, and etc.. However, this paper focuses on researches that use different types color information as a parameter for analysis and visualization of image data, such as hue, brightness and saturation. Our goal 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.

Our paper focuses on data collected from the 2013 Brazilian protests #vemprarua hashtag on Twitter, from the 15<sup>th</sup> of June to the 15<sup>th</sup> of July. The June 2013 Brazilian 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 sites gained great relevance, enabling protesters to rapidly share pictures, ideas, promote debate and events. It also enabled people at home to become part of this social movement, sharing and helping to promote and spread the news and events. Due to its importance to Brazil's social and political scenario, this paper also aims to better understand, through image analysis using color parameters, its contexts and repercussions

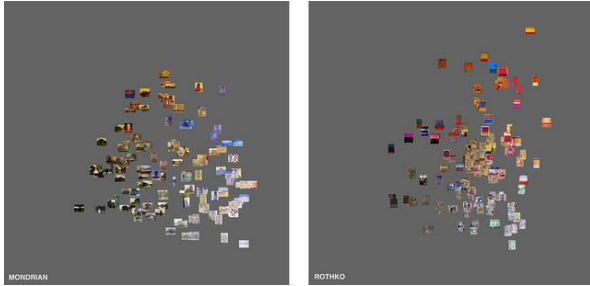
## 2. INTERNACIONAL STUDIES OF IMAGE BIG DATA ANALYSES BY COLOR

### 2.1 Color Analysis of Paintings

One of the spheres in which color analysis may reveal certain patterns is that of traditional work of art, such as paintings. Painters make use of a variety of colors to produce their works and define their artistic style. Using color parameters, we can perceive and analyze nuances that are present in works of different artists, enabling comparative analysis or even analyzes of what can be called as the stylistic “development” of a particular painter.

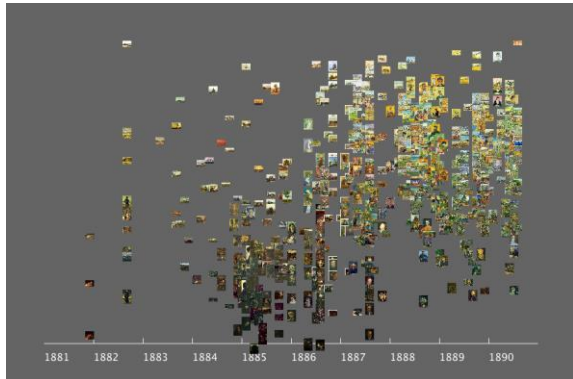
In June 2011, the Software Studies Initiative published a research analyzing two collections, one by Piet Mondrian and the other by Mark Rothko, based on the images' visual elements, such as hue, brightness and saturation, thus revealing patterns not only on the work themselves, but also the artists'. The purpose of this particular study was to compare a certain number of Mondrian's

paintings to Rothko's paintings, produced on similar periods of time in their career. This research identifies that, at the beginning of each time slot analyzed, their predominant artistic style was associated with the painting style used by their predecessors. But as the years went by, Mondrian and Rothko began to worry more about developing their own style of painting, trying to escape from figurativism.



**Figure 1a and 1b – Pictures of Mondrian (left) and Rothko (right). In each plot: X-axis (brightness medium) e Y-axis (saturation medium) [3]**

A different type of analysis can be done to associate color with time parameter, as in the study done with Van Gogh, analyzing his experience in Paris compared to when he moved to Arles. This research, also made by Software Studies Initiative, identifies that the set of images done in Arles contrasts with the set of paintings done in Paris, due to its higher saturation and brightness, a result of the painter's new color nuances experimentations. The visualization proposed by this lab (image2) aims to visually see how Van Gogh's painting was influenced by the city in which he found himself.

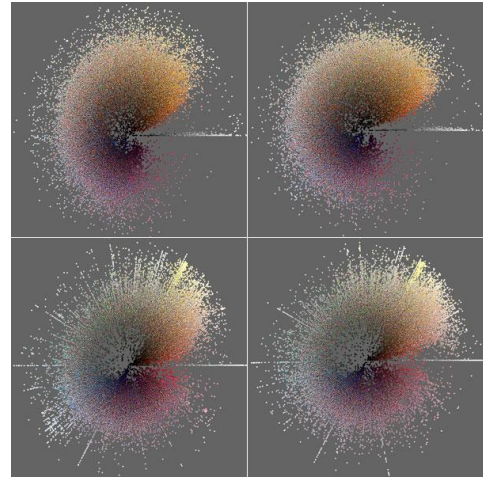


**Figure 2 - All images by Van Gogh, in Softwares Studies's dataset. X-axis: Month/Year Y-axis: brightness medium [3]**

## 2.2 Color Analysis of Phototrails

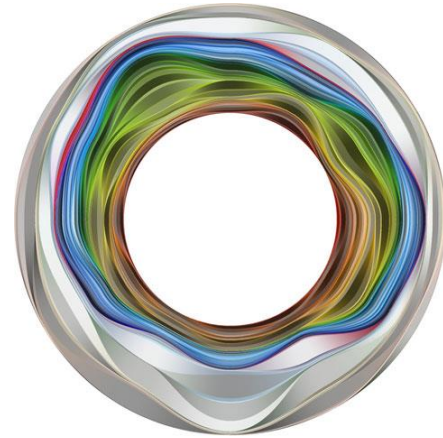
In July 2013, Nadav Hochman, Lev Manovich and Jay Chow, researchers from the Software Studies Initiative, developed a project called Phototrails. The proposed goal was to explore visual and dynamic patterns and structures on a planetary scale on photo sharing user-generated content. The study proposed to show, thorough analysis of visualizations of images from Instagram, how temporal changes and visual features of different locations can reveal social, cultural and political activities, as well as people's behavior 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 the datasets chosen, it was extracted basic visual information (such as average color, brightness, saturation, number of edges, contrast, etc.) to create different visualizations. The following images represents the visual identity of each city analyzed, taking into account the definitions made by these researchers:



**Figure 3 - NYC (upper left) and Bangkok (upper right) Radius: brightness medium Perimeter: hue medium [1]  
San Francisco (down left) and Tokio (down right) Radius: hue medium Perimeter: brightness medium [1]**

## 2.3 Color Analysis of Flickr Images - Flickr Flow



**Figure 4 - Time is disposed with summer in top and proceeding clockwise**

This is an example from 2009 of image visualization by color that uses photographs. The project was called Flickr Flow and developed by Fernanda Viégas and Martin Wattenberg, both data visualization researchers and explores. This study started with collections of photographs of Boston Common extracted from Flickr. With the available images' data, they divided all photos by month and calculated the relative proportions of different colors in each photo in each month of 2009. Its final step was to, using the parameters of colors and time, plot the colors on a wheel.

In this project, using color parameters, it became possible, within other results, to identify differences in climatic seasons of the year. At the bottom of this visualization, we will find a great amount of grays, whites and lighter colors, representing winter. We can then observe, clockwise, the increase of more vivid colors: variations of pink, purple, green and yellow - representing spring. Following this pattern, you can also see other seasons, such as fall indicated by yellow and orange, and summer by large amounts of bright colors, and a very few of white tones.

### 3. COLOR ANALYSIS IN HASHTAG

#### #VEMPRARUA

##### 3.1 The Brazilian Protests and the

##### #vemprarua Hashtag

The previous studies and researches conducted by the Software Studies Initiatives have shown the Laboratory of Studies on Image and Cyberculture, of Federal University of Espirito Santo (Brazil), the vast possibilities in working with huge amounts of images and in visualizing the behaviors and patterns that can result from social network users actions and contents. Through different types of tools and visualizations, our lab then proposes a study in which we can better understand the complexity and variety of political and social issues implied with the emergence of June's 2013 protests.

The object of this study, named "Visagem", is then to analyze twitter hashtag #vemprarua (which can be translated as "come to the street"), an iconic expression of the Brazilian protests and, thus, the most used hashtag to refer to this particular social movements within social media sites. 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 The Data Extraction

Our method for information capturing was datamining social networking service Twitter, through a software called yourTwrapperKeeper (a.k.a YTK), which uses Twitter API. With this method, all tweets that had the matching hashtag were extracted, creating a csv file with all the available information (such as who tweeted, date of publication, number of retweets, etc.).

Having in hands this csv file, we used a Java script called Crawler, created by our lab, whose function is to separate tweets that contain links and those that don't. Then this script goes into each tweeted link and captures the images that obey the parameters set previously by our researchers, such as a minimum size of 15 kB or 200 x 200 px, and extension file PNG, JPG, JPEG, TIF or TIFF.

In the period from June 15 to July 15, 2013, we extracted 85,595 images, coming from a total of 404,006 tweets. This was the period of big political protests in various Brazilian cities. These images, despite being posted on Twitter, come from a variety of websites, such as big news portals, online galleries, blogs, and other social networking sites.

##### 3.3 The Color Parameter of the #vemprarua Images

To analyze this large amount of images, we used HSB color scale, in which the color of each image pixel is composed by three numeric data (hue, brightness and saturation). Basically, the hue goes from 0 to 255 (equivalent to degrees 0° to 360°) forming a color circle. Brightness is determined by values ranging from 0 to 100, where zero means no light (dark) and 100 means the presence of bright (white). Finally, the variable saturation follows the numerical variation of brightness, also between 0-100, being 0 now an absence of tone - presence of gray - and 100 being fully saturated colors (bright, no gray). The chosen color scale basically helps identify groups of images with close measurements, and permits that we organize images by those parameters.

Afterwards, we used the plug-in "Measure", which is part of the software ImageJ, which can read the values of each image pixels and then calculates their hue, brightness and saturation. Thus, it was based on these three medians that this work could propose different visualizations and analysis

##### 3.4 The Analysis on #vemprarua 's Image Visualizations

With the images captured through YTK and with the visual information gathered by the Measure plugin, it was possible to plot visualizations in which large volumes of data are available. To make this type of plot, we used ImagerPlot, developed by Software Studies Initiatives housed within the UCSD Division of the California Institute for Telecommunication and Information Technology. From this plot, that can visually separates groups of images by its color parameters, an analysis was possible.

###### 3.4.1 Brightness x Saturation

Through the visualization below, we can see that the #vemprarua images are distributed among three major groups: a whiter set, mostly found in the upper left quadrant; a darker set, found at the base of the display; and a more colorful set, on the right upper quadrant.

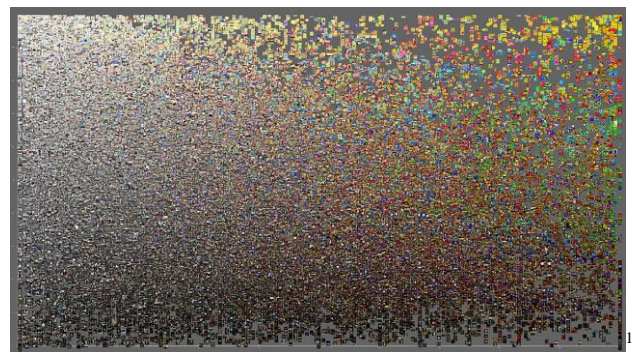


Figure 5 - 85.595 images sorted by X-axis (saturation median) and Y-axis (brightness median)

<sup>1</sup> Available in high definition on <http://zoom.it/G8jg>;

In the first group, with predominantly white images, we noticed a greater presence of posters, prints of documents and newspaper covers that have been shared throughout the month in which the protests occurred. The dissemination and circulation of information of this type was predominant throughout June 15 to July 15: such as posters shared as to incentive people's participation in the protests and to spread more information on the objectives and schedules of the events; documents and newspaper covers that contained information by the mass media; and others.

The second group consists of images of the exact moments in which protests occurred. They are images of a grayish tone resulting from the tone of the asphalt of the streets, visible during daytime protests, and darker, related to the nightly protests that have been the most striking feature of the event itself.

The third group is what unites the posters and advertisements broadcasted along with the #vempruarua. The posters in this group move away from the black and white pattern seeking more vivid colors, using mostly blue, green and yellow, referring to the national flag of Brazil.

### 3.4.2 Hue x Brightness and Hue x Saturation

When modifying the display parameters for "Hue" (X axis) and "saturation" (Y axis), the next visualization arranged the pictures in #vempruarua according to its color bands. Thereby, you can find groups of similar images and understand the frequency of appearance of certain types of images throughout the collection. The tracks that stand out are the warm colors (red, orange and yellow), green, blue, and the combination of purple and pink.

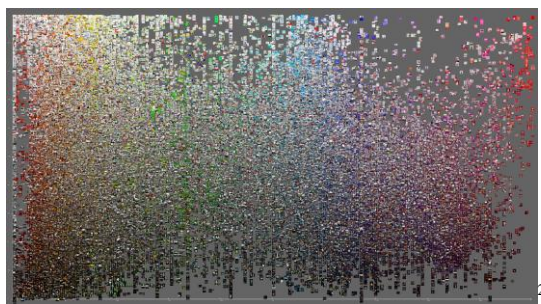


Figure 6 - 85.595 images sorted by X-axis (hue median) and Y-axis (brightness median)

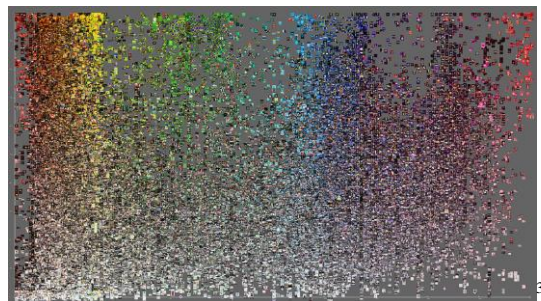


Figure 7 - 85.595 images sorted by X-axis (hue median) and Y-axis (saturation median)

The first color range is referred to warm colors and shows a largest number of images if compared with the rest. The

<sup>2</sup> Available in high definition on <http://zoom.it/QCYi>;

<sup>3</sup> Available in high definition on <http://zoom.it/tAaW>;

predominant theme that appears in this group is the protests in real time: photos taken at the time of the protests, although shared later. The predominant orange tone is related with the tone of street lighting that accompanied the acts performed in the evening; and also photos of confrontation, which appear fires, explosions and shots fired by police.

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 manifestation posters. The white band that bears the inscription "Order and Progress" (Ordem e Progresso) was replaced, for example, "In Progress" (Em Progresso), meaning that the country was in a state of change led by the people.

The blue color range is composed of images of flags of Brazil focusing on the inner circle of the design, and a lot of photos of Instagram, since one of the available filters to be applied in darker photos gives a bluish tint to them. The last color range covering the pink and purplish tones are pictures of the mobilization that were a bit faded and posters with a message tone calmer and soft, almost feminine.

### 3.4.3 Just Color Visualization Order by Hue with Static Brightness and Saturation



Figure 8a and 8b - Visualization using the median values of hue, saturation and brightness

After creating the previous plots using ImageJ, we proposed a different kind of visualization. Instead of placing an image in a specific position determined by its color parameters, in the image above we placed just the median colors of the images. Thus, in the end, it showed a graphic with the median colors of each image.

For this visualization, a script was developed in our lab using Processing language to visualize the plot of images from the medians colors calculated by ImageJ, in which each image is represented by a square of 2 by 2 pixels. Through the hue median (0 – 255), the squares representing each image were plotted in the ImageJ's visualization. Thus, the top image represents the images with hue median 0 and bottom with hue median 255.

We used two types of sets of color parameters that resulted in two different graphs. For the first graph, we used the hue median, saturation median and brightness median, as to compose the color

of the squares representing the images. For the second graph, we used just the hue median of the image; the saturation and brightness were established in a standard value. So in the first graph we have a visualization of all the colors parameters of all images, and in the second graph it's possible to see more clearly just the hue value.

#### 4. CONCLUSION

Based on what has been presented throughout this paper, we can understand the importance of the usage of the color parameters for visualizing large amounts of images. Both in the artistic field and in the studies of social movements, we see that color value can reveal more than numeric information.

When organizing a large volume of images by color, brightness and saturation parameters, we can see that certain groups are formed, which may reveal behavioral patterns of the network that is being analyzed. In the visualizations created with ImageJ, we can see the distribution of the images in the graph and notice that the largest amount of shares was dark and yellowish images. This may be related to an increased frequency of nocturnal presence in protests, whose numbers reached the houses of hundreds of thousands in some metropolitan areas. Another pattern to highlight is the formation of groups of similar images within the color display that has stripes: we are then able to identify that the green band is represented by a group of flags of Brazil, or the blue band has predominantly selfies of Instagram, serves to support the hypothesis that the analysis of the visual parameters of the images goes beyond esthetic and reveals the behavioral patterns of the events.

However, the visualizations made with ImageJ have certain limitations that hinder a deeper analysis of the datasets. The plots of the images are made using only two coordinates (X and Y), so when a picture has the same coordinates as another image, there is an overlapping and you lose image information. In view of this problem, we perceive a need to create a tool capable of adding a Z axis allowing a 3D environment, so that the information from the images that are positioned behind would not be lost.

When using Processing, each image is represented by its corresponding pixel, with no overlap. In the plot range of colors (topic 3.4.4 - left) you can see the width proportion of each color range in the dataset, confirming the theory, that arose in the visualizations made on ImageJ, that mostly orange toned pictures were shared during the protests. The right plot (topic 3.4.4 - right) follows the same principles of the plot on the left, in which the

image is represented by pixel color values, brightness and saturation (HSB), and that shows that despite the predominant color being orange, the prevalence of tone of the protests is dark, which again refers to the striking feature of the June demonstrations which took place in the evening.

So this paper concludes 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 behavioral patterns. When people paint or share a certain type of image, it can associated with the environment where they are and with the thoughts and feelings they have. When we analyses image visualizations, we end up seeing beyond numerical values, and seeing the various angles of an determined fact. Thus, the color value enables us to create numerous types of plots to visualize the data and extracted varied information that is found disorganized amid raw data.

#### 5. ACKNOWLEDGMENTS

Our thanks to the Federal University of Espírito Santo (UFES), National Council for Scientific and Technological Development (CNPq) and the Foundation of Suport for Research at Espírito Santo (FAPES) for the financial suport of the research. This research is a part of the "Visagem" project, carried out by the Laboratory of Image and Cyberculture Studies (Labic).

#### 6. REFERENCES

- [1] Hochman, N., Manovich, L., Chow, J. 2014. *Phototrails*. Available on: < <http://phototrails.net/>>.
- [2] Manovich, L. *Data Visualization and Computational Art History*. 2012. Available on: <<http://lab.softwarestudies.com/2012/04/data-visualization-and-computational.html>>
- [3] MANOVICH, Lev. *Style Space: How to compare image sets and follow their evolution*. 2011. Available on: < <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 on <<http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698>>