Weather Condition Effect on Webcam Image colorization

Ben Tucker

Abstract

Weather conditions have a significant effect on the colorization of webcam images. By using simple categorical weather condition data provided by the National Weather Service, models can be constructed to account for up to 24% of the variation in levels of red and 22% of the variation in levels of blue.

Introduction

Webcams have become increasingly pervasive in public places. Their intended uses can range from crime prevention to traffic monitoring to virtual tourism. Many of these cameras are open for the public to download a stream of either static images or full motion video. This allows for webcams to be used for purposes beyond their intended utility. The privacy implications of webcams being leveraged for invasive surveillance are real, as the images can be automatically analyzed to recognize cars, groups of people, and even faces. Such automated analysis is made more challenging by the fact that the colorization of images can shift considerably due to a changes in sunlight, precipitation, and other weather effects.

This study aims to show that publicly available weather data for the vicinity of the webcam can be used to better account for some of the variability in image color.

Methods

For this study 26 webcams were used that are operated by the Pennsylvania Department of Transportation[1]. All webcams were located in the vicinity of Pittsburgh, PA and update several times per minute. Weather information was retrieved from the National Weather Service's Experimental XML data feeds using data originating at Pittsburgh International Airport[2].

A Perl script was run every 20 minutes which first downloaded the current weather conditions, and then downloaded the current image from each of the 26 cameras. Information was stored in a MySQL database.

At the end of a four day period, four different times were chosen to analyze across the four days: 9am, 12pm, 3pm, and 6pm. From these 3pm was focused on for the remainder of the analysis, due to its greatest variation of weather over study period. Data collection continued for another 10 days.

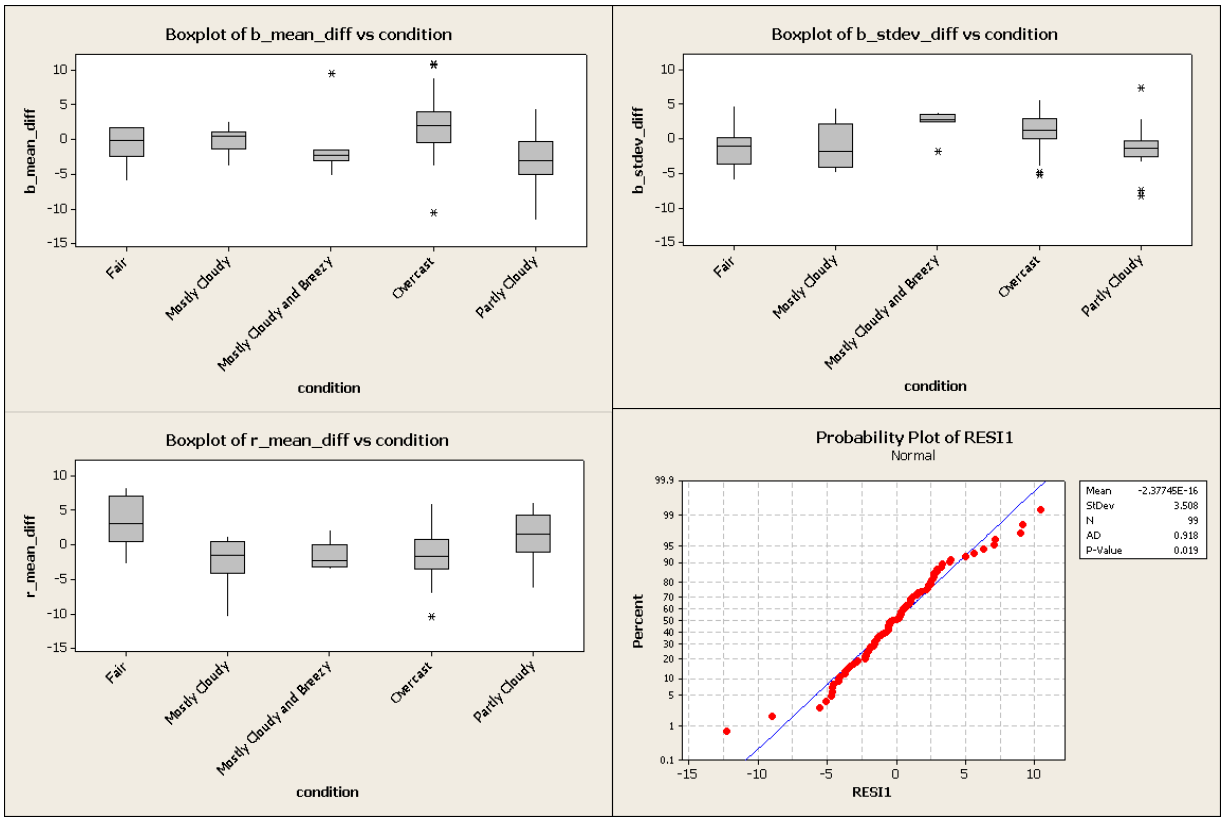
Unfortunately, the webcam operators have controls to move the cameras to a variety of vantage points. This hinders the ability to compare images over time. To resolve this, all 3pm images were individually examined to determine if the camera angle had changed significantly from other frames taken by that webcam. All days which did not match were discarded. Webcams whose angles constantly changed were discarded entirely, leaving only 10 of the original 26 cameras and 99 images in the study.

1 http://www.nb.net/~finals/allcams.htm
2 http://www.weather.gov/data/current_obs/KPIT.xml

For each image remaining, the jpeg taken 20 minutes before and after were averaged into the 3pm shot to help reduce the effects of non-weather related artifacts such as traffic. Each of these composite images were then fed through ImageMagick[3] to determine the mean and standard deviation of each of the red, green, and blue channels. Then, to normalize for the different camera views, a "difference from mean" was computed for each image based on the average of the respective camera's standard deviation and mean for each channel. The equations used were based on the following example: b_mean_diff = blue_mean – AVG(blue_mean for all images from this camera).

This data was exported from the database and imported into Minitab[4] for statistical analysis. First some exploratory analysis was done to determine what correlations appeared to exist in the data. In addition to categorical weather condition information, numerical weather data was also examined. This data included dewpoint, humidity, barometric pressure, and temperature. However, no significant correlations could be found between any of these indicators and the image colorization. Thus, only the categorical condition ("Fair", "Partly Cloudy", "Mostly Cloudy", "Mostly Cloudy and Breezy", "Overcast") was used in the analysis.

Figures 1.1 – 1.3 show the boxplot graphs of the variables which exhibited the greatest evidence of a correlation. An ANOVA calculation was then done to fix each variable independently to the model and determine the significance. Figure 1.4 shows that the residuals were not normally distributed, meaning there are other factors at work not accounted for by the model and meaning that we must be cautious with the conclusions we draw from the data.



Figures (clockwise from top left) 1.1, 1.2, 1.3, 1.4

3 ImageMagick is an image processing library - http://www.imagemagick.org
4 Minitab is a statistical analysis software package - http://www.minitab.com

Results

There is a significant relationship between weather condition and the following variables: blue standard deviation, blue mean, and red mean. All other tested variables did not result in significant relationships.

Figure 1.5 shows the results of a One-way ANOVA with condition as the explanatory and blue mean as the response variable. It shows us that the the model can explain 22.21% of the variation in the difference between the mean of an image's blue channel and the average of the means of all other images from that camera. It also shows us that there is the greatest amount of blue in the "Partly Cloudy" condition. The "Overcast" condition has the least amount of blue. See figure 1.9 for an example. The "Fair" condition has less blue than "Partly Cloudy" because without clouds in the sky, the camera has a difficult time judging contrast and tends to display the entire image lighter (and with greater shades of gray).

One-way ANOVA: b_mean_diff versus condition

```
Source      DF      SS    MS     F      P
condition    4   344.2  86.1  6.71  0.000
Error       94  1205.8  12.8
Total       98  1550.0

S = 3.582   R-Sq = 22.21%   R-Sq(adj) = 18.90%
                                    Individual 95% CIs For Mean Based on
                                    Pooled StDev
Level              N    Mean  StDev  ----+---------+---------+---------+-----
Fair              10   0.889  2.779         (----------*----------)
Mostly Cloudy      9   0.080  1.875     (----------*----------)
Mostly Cloudy an   7   1.089  4.785        (------------*-------------)
Overcast          47  -1.699  3.708  (-----*----)
Partly Cloudy     26   2.724  3.679                       (------*------)
                                    ----+---------+---------+---------+-----
                                      -2.0       0.0       2.0       4.0
```

Figure 1.5

Figure 1.7 shows the analysis of the standard deviation of the blue channel. From it we can see there is significantly less variability in the overcast condition than any other condition with the exception of "Mostly Cloudy and Breezy". This latter condition only occurred on one day, so the data is limited and could be considered an outlier.

Figure 1.8 shows the analysis of the mean of the red channel. This model proves most effective with an R-sq value of 23.86%. We see that in the "Fair" condition there is significantly less red on average than in all over conditions. This is exemplified in Figure 1.6

Figure 1.6: Left is "Fair" condition, right is "Overcast"

One-way ANOVA: b_stdev_diff versus condition

```
Source     DF      SS     MS     F      P
condition   4  172.27  43.07  5.96  0.000
Error      94  678.75   7.22
Total      98  851.01

S = 2.687   R-Sq = 20.24%   R-Sq(adj) = 16.85%

                                    Individual 95% CIs For Mean Based on
                                    Pooled StDev
Level              N    Mean  StDev  --+---------+---------+---------+-------
Fair              10   1.365  3.015                     (--------*-------)
Mostly Cloudy      9   1.174  3.322                    (--------*--------)
Mostly Cloudy an   7  -2.373  1.900   (---------*---------)
Overcast          47  -1.122  2.434              (--*---)
Partly Cloudy     26   1.301  2.929                        (-----*----)
                                     --+---------+---------+---------+-------
                                    -4.0      -2.0       0.0       2.0
```

Figure 1.7

One-way ANOVA: r_mean_diff versus condition

```
Source     DF      SS    MS     F      P
condition   4   313.2  78.3  7.37  0.000
Error      94   999.2  10.6
Total      98  1312.4

S = 3.260   R-Sq = 23.86%   R-Sq(adj) = 20.62%

                                    Individual 95% CIs For Mean Based on
                                    Pooled StDev
Level              N    Mean  StDev --------+---------+---------+---------+-
Fair              10  -3.437  3.785 (------*-----)
Mostly Cloudy      9   2.384  3.642                   (------*------)
Mostly Cloudy an   7   1.585  1.992                (-------*-------)
Overcast          47   1.327  3.209                   (--*---)
Partly Cloudy     26  -1.382  3.266          (---*----)
                                    --------+---------+---------+---------+-
                                         -3.0       0.0       3.0       6.0
```

Figure 1.8

Figure 1.9

Discussion

While the three significant ANOVA models do show there is a significant relationship between the weather condition and the red/blue colorization of the webcam images, the fact that the residuals, upon applying the model, are not normally distributed means what is being observed could be the result of other factors, or flaws in the study. Problems could include that the number of observed instances of each weather condition are far ranging with “Mostly Cloudy and Breezy” being the condition in only 7 images, while “Overcast” is the state of 47 images. A possible solution to this issue is to combine weather conditions into just two groups: “Fair & Partly Cloudy” and “Mostly Cloudy (and Breezy) & Overcast.”

Additionally, it may be that each webcam needs to be analyzed independently. This could lead to considerably more accurate models. It's also possible that logistic regression could be more effective than ANOVA.

Conclusion

From the data collected it is clear there are differences in the colorization of webcam images during different weather concisions. The statistical analysis shows this relationship is significant. The red channel is greater in cloudy and overcast concisions, while the blue channel is much stronger in Fair and Partly cloudy weather. This is consistent with visual observations from the data. While our model is only able to explain at most 23% of variation, this is still a significant relationship and shows that this method could be put to use in other models to better improve image analysis.