ART STATISTICS AND VISUAL PROCESSING: INSIGHTS FOR PICTURE CODING

Daniel Graham

Department of Mathematics, Dartmouth College

ABSTRACT

Artwork holds much important information regarding the efficient encoding of the natural world, and it is therefore useful both for researchers in vision science and those in signal processing. Painters, like photographers, aim to capture the visual environment in a way that is appealing to viewers. But until recently, little attention has been paid to statistical regularities related to artists' representational strategies. How do artists deal with the very large dynamic range of luminances in scenes, when paintings themselves have a far smaller dynamic range? To what extent do artists reproduce the scale invariant spatial statistics of natural scenes, and what statistical regularities of natural scenes, if any, are retained when artists paint abstractly? This paper discusses findings that shed light on these questions and it suggests ways that these findings could spawn novel strategies for picture coding and image retrieval. It also describes links between artists' representational strategies and neural coding in visual systems.

Index Terms—vision, retina, HDR, art, luminance scaling, gamma, CBIR, efficient coding, perception, similarity metrics.

1. INTRODUCTION

This paper discusses novel approaches to luminance and spatial coding, as well as to image retrieval, gleaned from examining the ways fine art painters mimic and modify the statistics typical of the natural world. One would expect that since painters create works that are intended to be seen by other humans—and since the medium of paint is quite limited in its ability to capture the full statistics of the natural world—that methods employed by painters could be of much use to the picture coding community. Indeed, unlike photographers, painters have the ability to manipulate every point in their representation. What then are artists' shared strategies for representing the natural world?

Work in this vein is mostly directed toward furthering our understanding of coding in the visual system, though its applicability to picture coding will be the focus of the present paper. With regard to the visual system, the study of natural scene statistical regularities (spatial frequency, color, luminance, etc.) can be used to infer coding strategies in the early visual system by means of the efficient coding hypothesis. This hypothesis proposes that shared statistics in scenes will be "assumed" to be present by visual system encoding mechanisms in such a way as to efficiently use metabolic, representational, and learning resources. In other words, the visual system will have adapted over evolutionary time to the redundancies typical of a creature's natural environment, and thereby make the task of seeing easier. We see much supportive evidence for this notion in retinal coding and early cortical coding. Moreover, many of the implicated coding strategies are ones shared by many if not most sighted animals.

This guiding principle can be extended to the realm of artwork, which we have argued is special because of the motivation of artists to produce works for human viewing. We can suppose that art is typical of the human visual environment, and therefore that it contains statistical regularities above and beyond those found in natural scenes. To the extent that such regularities exist, arguments along the lines of the efficient coding hypothesis can be employed, and to greater possible effect. There is the possibility of learning about poorly understood aspects of human vision in particular, since we can view art as the environment most typical for humans. For example, as discussed below, the encoding of high-level perceptual properties, like style and similarity, shows evidence of the influence of statistical regularities. In addition, paintings are plentiful: they have been produced for over 30,000 years and are now widely available in digitized form. This approach holds special promise for understanding differences between the human visual system and that of the macaque, whose visual system is studied as a model of our own.

The following section begins with a survey of regularities in luminance and spatial statistics found in artwork from around the world and across history. Next, we describe experiments that demonstrate that variations in these statistics can be good predictors of perceptual properties. This latter finding has implications both vision scientists' understanding of the encoding of basic perceptual dimensions such as similarity and it also provides insights for image search applications. We conclude by sketching out future directions for this research area, which will include investigations of regularities in color statistics in painted artwork and temporal statistics in popular film.

2. ART AND NATURAL SCENE STATISTICS

Apart from optics, the chief technology that permits imaging is compression, i.e., the mapping of the intensities of the natural world to the limited sensitivity range available in photographic film or electronic sensors. And while optics are largely a matter of physics, effective compression depends greatly on biological factors, and specifically on the response of the visual system. It follows that a thorough understanding of the statistical regularities in the natural world—luminance, spatial frequency, color, edge co-occurrence, etc.—is necessary if one hopes to produce an effective imaging system, and also if one seeks to understand neural coding in the visual system.

It is interesting that the characterization of the statistical regularities in the natural world was initially investigated by imaging researchers. In 1949, Jones and Condit provided the definitive study of natural luminances, while in 1952 Kretzmer reported findings about spatial statistics of television signals. However, until recently, the bulk of the empirical work in this area has been carried out by vision researchers investigating efficient coding. It is also intriguing to find that engineers are aware that art holds a wealth of information about natural scene representations [1], yet few if any large scale studies of art statistics have been carried out in any field until recently.

With regard to vision, consider that though the space of possible images is vast, most such images (i.e., white noise) are equally imperceptible. This can be seen as evidence that supports an intriguing idea: the visual system did not evolve to distinguish noise images because the world does not resemble noise. A theoretical visual system *could* become attuned to the large statistical differences among white noise images, and thereby distinguish different classes of them. But why waste valuable coding units (i.e., neurons and networks of neurons) to process images that never occur in nature? Moreover, there is evidence that the 1/f fall-off in spatial frequency spectra typical of natural scenes is matched by the spatial frequency sensitivity of retinal ganglion cells: these cells appear to produce a decorrelated and/or equalized response [2] to scenes. This finding suggests the visual system has achieved an efficient (redundancy reduced) representation.

2.1. Spatial statistics of artwork

As one might expect, the basic 1/f spatial frequency amplitude spectrum typical of natural scenes is replicated in art. A study of 140 paintings spanning 900 years and both hemispheres found that art has a 1/f shaped amplitude spectrum, but that the mean fall-off of the amplitude spectrum for the artworks was significantly less than that of a collection of calibrated natural scenes (measured as slope of the spectrum plotted on log-log axes; for natural scenes, the slope was –1.4, for art it was –1.2) [3]. It is unclear why the fall-off is different: it may reflect the fact that Eastern Hemisphere works—which had significantly shallower spectra than Western works—comprised a large proportion of the sample. In such works, global structure is more uniform (less low frequency power from, e.g., strong horizon lines) but fine detail is emphasized.

It should be noted that a separate study by Redies et al. [4] showed no significant difference between the mean amplitude spectrum slopes for Western graphic art (drawing, engraving) and natural scenes. This team also showed that for art, this statistic is not sensitive to changes in gamma during imaging.

A result that may be more surprising than the finding of similar spatial frequency spectra for art and scenes relates to abstract art. Whereas many abstract works are created with some degree of randomness (e.g., Jackson Pollock,), they nevertheless show 1/f fall-off in their spatial frequency spectra. Some artists do manage to escape this regularity by painting spatially periodic patterns, or monochromes, though 1/f-shaped amplitude spectra appear to be the norm.

However, the mean slope for a collection of works judged by observers in a three alternative forced-choice test to be "abstract" was significantly shallower than the mean for images judged to be representational (i.e., for those judged to be "landscape" and "portrait/still-life") [5]. This difference may relate to the lack of shading in abstract art, which serves to soften edges and steepen spectra. Shading is not required in most abstract works because there is rarely implied lighting, nor are there generally any objects. Further tests are needed to validate this hypothesis.

Individual artists, as well as defined periods in an artist's oeuvre also appear to have distinctive spatial statistics. Lyu and colleagues [6] used feature vectors composed of wavelet decomposition coefficients to separate drawings by Pieter Bruegel the Elder from those of Bruegel's known imitators. Related work has used sparse coding models—which were first developed to study efficient coding in the primate visual cortex—to predict the likely historical dating of three Van Gogh paintings whose production date is debated by art historians [7]. The sparse coding method is particularly well suited to the task of stylometric classification because (1.) the basis functions of the representation are learned by a neural network architecture from image samples, rather than being pre-specified as with, say, wavelets; (2.) the statistical features to which sparse coding is attuned have a structure similar to brushwork in the paintings, which is a primary factor in art historical analysis for assigning an undated work to a given historical period in an artist's life. Hughes et al. find good agreement on the likely classification at three different spatial scales. They also replicated the Bruegel discrimination result of Lyu et al. This shows that the sparse coding method is robust to different types of art and is sensitive to distinctive styles used by an artist and his imitators, and even in the work of the same artist. Therefore, picture coding algorithms that are sensitive to sparse and higher-order statistical structure may be especially useful for artwork, and for images more generally.

2.2. Luminance statistics of artwork

Using the calibrated image sets described in an earlier paper [6], we have found that natural scene luminance distributions are highly skewed on average (mean skewness of 3.28, standard error: 0.27), while art images distributions have no skewness. This regularity affects the kurtosis (a measure of sparseness) of these distributions, which is again far higher for natural scenes (mean for natural scenes: 31.3; for art: 1.0). As noted earlier, effective luminance compression is necessary to produce visible images of natural scenes. Given the highly non-Gaussian shape of the typical luminance distribution for scenes—and the nearly Gaussian shape found for art—an effective transform between the two must be nonlinear.

One can model the effect of luminance compression to a first approximation by using the well-known technique of histogram matching: the function that transforms the scene luminance histogram into the painting luminance histogram can be viewed as a gamma function or look up table. For paintings, this is called the artist's look-up table or ALUT [11]. This model requires that a suitable scene can be found to match the original tableau, which is difficult for older works. There have been no large-scale studies of multiple artists measuring the artist's look-up table but studies of individual artists have been completed. In general, this work has shown that artists (including Van Gogh [8]) do indeed use nonlinear ALUT tone mappings [9]. Van Gogh in particular uses sigmoidal ALUTs, as shown in Figure 1. This figure refers to a tableau in Arles, France that corresponds to the Van Gogh painting designated as F412 ("Harvest Landscape," 1888). The figure shows the modern tableau of F412 matched to the luminance histogram of the F412 painting and to the histogram of another Arles work, F486, "Les Alyscamps" (1888). Though the ALUTs differ only slightly, their effect is noticeably different. Notice that the shadows under the near-ground bushes are detailed, but in the image mapped with the non-matching transform (F486), they are dark.

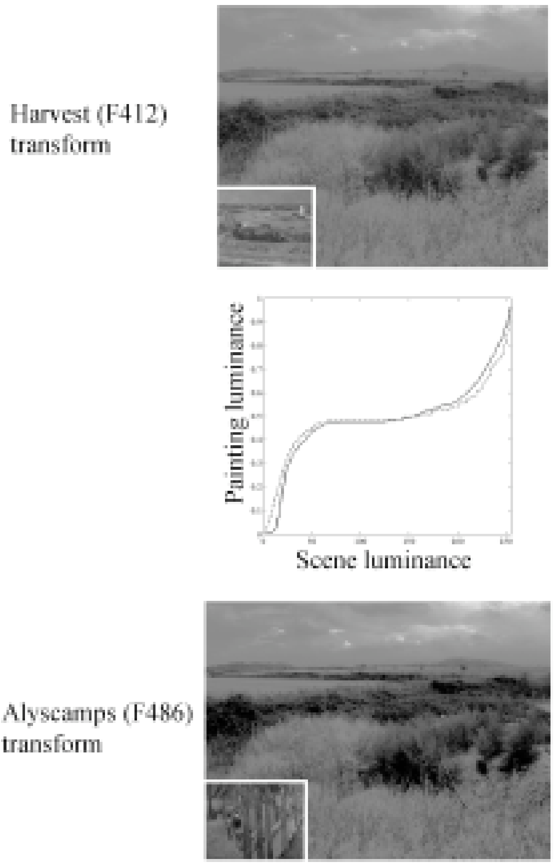

Figure 1. On top, scene transformed (tone mapped) to F412 "Harvest Landscape," (inset). This transform is shown as the dotted line in the center plot. Solid line is the ALUT transforming the same scene to F486 ("Les Alyscamps"), and the result of this transformation is on the bottom, with F486 inset. In center plot, horizontal axis represents scene luminances (linearly scaled to 0:255), vertical axis represents painting luminances (linearly scaled to 0:1). F412 and F486 images © Van Gogh Museum, Amsterdam.

The consistency of the two ALUTs also suggests that Van Gogh's general strategy of using sigmoidal scaling is robust for two images with very different illumination: F412 depicts an open scene with sky, while F486 depicts a forested scene with no sky. That is, Van Gogh may have attempted to achieve consistent luminance transformations in his paintings despite changing luminance conditions in depicted scenes.

Note that the ALUT for F412 can be inverted (so as to transform painting luminances to world luminances), which gives a familiar sigmoidal shape. Sigmoidal tone mapping functions are used often in imaging when detail in dark areas needs to be preserved. The inverted ALUT can be thought of as the global nonlinear luminance stretching that would be applied to a painting in order to scale it to match corresponding scene luminances. This has strong parallels in the high dynamic range (HDR) graphics literature. For instance, one group [10] has used a scheme involving a sigmoidal-shaped look-up table (along with retina-like spatial filtering) in order to simulate "painting" HDR images using a low dynamic range display. Van Gogh—and other painters—would appear to employ a representational approach that mirrors this modern HDR tone mapping strategy.

Note that no spatial information is used in this model. Local luminance adjustments by painters are surely important but are not captured by the ALUT. Nonetheless, this method indicates that tone mappings of the same general form as the Van Gogh ALUT could be good all-purpose transforms.

3. ART STATISTICS AND PERCEPTION

Much evidence supports the notion that basic statistical regularities (e.g., 2-D amplitude spectra) that are "diagnostic" of scene categories could in principle be extracted by the early visual system. Studies of categorization for natural images have demonstrated relationships between basic image statistics and classes of natural scenes [11], and between basic statistics and perceptual judgments [12]. These same statistics, (e.g., skewness), are relevant to efficient visual coding in primates. It is not yet clear whether variations in these statistics are in fact used by the visual system to determine scene class, context, or other perceptual information. But in any case, this predictability can be built into picture coding models.

Recent studies investigated whether basic statistics are related to variations in perceived similarity of artwork, [8,13]. These studies employed multidimensional scaling analysis of observers' similarity ratings for paired paintings, and resulting dimensions were compared to a host of statistical measures, modeled neural responses, and semantic variables derived from image metadata. A large proportion of variance was explained by a statistic called the activity fraction. The activity fraction S over n pixels each with intensity r_i has a range of 0 to 1:

$$S = \frac{\left(\frac{1}{n}\sum_{i}^{n} r_i\right)^2}{\frac{1}{n}\sum_{i}^{n} r_i^2}$$

This nonlinear measure is typically used to gauge the sparseness of neural population responses, though it can be applied to any non-negative data. Small values of S (near zero) correspond to a highly sparse, heavy-tailed distribution of intensities, where a few pixels have high intensities and the rest show low intensity. For both landscapes and portraits, one of the first two similarity dimensions was highly correlated with the activity fraction, and for landscapes this correlation explained a greater portion of data variance than did semantic variables (such variables encode the presence or absence of humans in a landscape, for example). Spatial statistics such as spatial frequency amplitude spectrum slope were nearly as good predictors compared to semantic variables. Given these results, we have turned our attention to abstract art using the same procedure. Abstract works are especially interesting because they generally lack semantic content. We found that a sum and a difference of the activity fraction measure of sparseness and the log-log slope of the spatial frequency amplitude spectrum explain roughly a third of the overall variance in the similarity data. The summing and subtracting operations serve to rotate the image data so as to better align with empirical dimensions (i.e., the first two MDS scales, which together explain 45% of similarity data variance). MDS scale 1 and the summed statistics have a correlation of .85 ($p < 0.0001$); MDS scale 2 and the subtracted statistics have a correlation of .81 ($p < 0.0001$). A linear fit of the two relationships explained 72% and 66% of the data variance, respectively. That is, together these two statistical measures explain nearly a third of the overall variance in perceived similarity ratings of human observers. This result demonstrates the utility of higher-order and nonlinear statistical features in assessing perceptual dimensions. As noted earlier, higher-order statistical properties such as skew and kurtosis (as well as activity fraction) are starkly different for scenes and for art.

This work suggests that such statistics are useful more generally in image search and content-based image retrieval (CBIR). Future studies using artworks will test a model that predicts similarity judgments based on image statistics against a number of standard tag-based techniques. The goal is to see how far one can go in predicting viewer judgments of similarity (and the related judgment of preference) using spatial, luminance and color statistics alone. As has been argued, art images are generally restricted in a statistical and semantic sense compared to natural scenes. This fact may make statistics-based CBIR algorithms that measure nonlinear image statistics effective when applied to art databases, and this may be especially so within agreed-upon categories of artworks.

Indeed, there is increasing need for fast and accurate search algorithms for art and film databases. For example, though the highly successful Netflix, Apple Genius, and Pandora programs all use predictive search algorithms, none employs statistical measures of the movies or music in their respective databases. Perhaps art will prove to be different—that is, it may prove effective to match image statistics in order to perform searches within art databases. Art, unlike music or film, can be instantly evaluated. Therefore, gross differences in luminance, color, and spatial statistics could be the overriding factors in basic perceptual judgments of images, at least for artworks, as the studies cited above demonstrate. It may be the case that artists signify basic semantic or perceptual data in part using predictable variation in image statistics, and perhaps they do so using a strategy similar to the one employed by the visual system. CBIR algorithms could exploit such predictability, even if the predictability is not ultimately related to the manner in which the visual system encodes perceptual dimensions.

Of course, user-generated content tags (e.g., to assign a work general content categories, such as "landscape") will also be important in such search algorithms. But tags have the disadvantage of requiring input from one or more users prior to the search. Relevant image statistics, on the other hand, can be calculated on the fly, and for novel images.

4. FUTURE DIRECTIONS

Color statistics have been found to play a large role in behavioral studies of image perception, and they can be used for image search. Color statistics have also been studied in relation to optimal lighting of artwork. It follows that color statistics are the next frontier, and that large-scale study of statistical regularity in art is needed. This is an especially promising direction given the highly systematic nature of color application in art, which includes both the mixing of pigments and the practical theory for spatial arrangements of complementary colors. Of course, this aspect of the artist's representational strategy, like the others, is dependent on the basic coding properties of the human visual system. However, less is known about color coding in the visual system, compared to spatial and luminance coding. What is typical in terms of perceptually relevant color statistics (e.g., CIELAB dimensions), and color opponent statistics? How do these statistics compare to color statistics of natural scenes? Answers to these questions will no doubt be of use to both the vision community and the signal engineering community.

Video is another frontier: a large, multi-institutional effort is beginning to study statistics of large sets of digitized artwork, and one branch of this project will examine statistical properties of popular film from its creation through today.

5. CONCLUSION

This paper has summarized the body of recent research devoted to the characterization of statistical regularities in artwork. We have shown that regularities found in art are particularly germane to the future development of picture coding strategies. In particular, it has been suggested that painters use efficient nonlinear tone mapping strategies, which could be exploited by coding algorithms. Also, nonlinear statistical regularities within classes of artworks (both spatial and luminance statistics) are surprisingly good predictors of basic perceptual judgments, a fact that could inform new CBIR strategies.

Acknowledgements: This work was supported by National Science Foundation grant DMS-0746667 and by the William H. Neukom 1964 Institute for Computational Science.

7. REFERENCES

[1] J. Tumblin, J.K Hodgins, and B.K. Guenter, "Two methods for display of high contrast images," *ACM Trans. Graph*. **18** 56–94, 1999.

[2] D.J. Graham, D.M. Chandler and D.J. Field. "Can the theory of whitening explain center-surround properties of retinal ganglion cell receptive fields?" *Vis. Res*. **46**, 2901-2913, 2006.

[3] D. J. Graham and D. J. Field, "Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities," *Spatial Vision* **21**, 149-164, 2007.

[4] C. Redies, J. Hasenstein, and J. Denzler, "Fractal-like image statistics in visual art: similarity to natural scenes," *Spatial Vision* **21**, 137-148, 2007.

[5] D. J. Graham and D. J. Field, "Variations in intensity statistics for representational and abstract art, and for art from the eastern and western hemispheres," *Perception* **37**, 1341-1352, 2008.

[6] S. Lyu, D.N. Rockmore and H. Farid, "A digital technique for art authentication," *Proc. Nat. Acad. Sci. USA* **101**, 17006-17010, 2004.

[7] J.M. Hughes, D.J. Graham and D.N. Rockmore, "Stylometric classification for Van Gogh paintings of undetermined provenance using sparse coding," In prep. for *Patt. Recog*.

[8] D.J. Graham, J.D. Friedenberg, D.N. Rockmore, "Efficient visual system processing of spatial and luminance statistics in representational and non-representational art," *Proc. SPIE* **7240**, 2009

[9] D. J. Graham and D. J. Field, "Global nonlinear compression of natural luminances in painted art," *Proc. SPIE* **6810**, 2008.

[10] M. Colbert, E. Reinhard and C. E. Hughes, "Painting in high dynamic range," *J. Visual Comm. and Image Repr.* **18**, 387-396, 2007.

[11] A. Torralba. and A. Oliva, "Statistics of natural image categories," *Netw.: Comput. Neural Syst.* **14**, 391-412, 2003.

[12] I. Motoyoshi, S. Nishidam, L. Sharan. and E. H. Adelson, "Image statistics and the perception of surface qualities," *Nature* **447**, pp. 206-209, 2007.

[13] D. J. Graham, J. D. Friedenberg, D. N. Rockmore and D. J. Field, "Mapping the similarity space of paintings: image statistics and visual perception," *Visual Cognition*, in press.