Does Color Influence Image Complexity Perception?

Gianluigi Ciocca[1,3], Silvia Corchs[1,3], Francesca Gasparini[1,3(✉)], Emanuela Bricolo[2,3], and Riccardo Tebano[2]

[1] Dipartimento di Informatica, Sistemistica e Comunicazione, University of Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy
{ciocca,corchs,gasparini}@disco.unimib.it
[2] Department of Psychology, University of Milano-Bicocca, Via Dell' Innovazione 10, 20126 Milano, Italy
emanuela.bricolo@unimib.it, r.tebano@campus.unimib.it
[3] NeuroMi - Milan Center for Neuroscience, Milan, Italy

Abstract. In this paper we investigate if color influences the perception of image complexity. To this end we perform two different types of psycho-physical experiments on color and grayscale images. In the first experiment, images are ranked based on their complexity (image ranking), while in the second experiment the complexity of each image is assessed on a continuous scale (image scaling). Moreover, we investigate if ten image features, that measure colors as well as other spatial properties of the images, correlate with the collected subjective data. The performance of these correlations are evaluated in terms of Pearson correlation coefficients and Spearman rank-order correlation coefficients. We observe that for each type of experiment, subjective scores for color images are highly correlated with those of the corresponding grayscale versions suggesting that color is not a relevant attribute in evaluating image complexity. Moreover none of the tested simple image features seem to be adapt to predict the image complexity according to the human judgments.

Keywords: Image complexity · Psycho-physical experiment · Color image features

1 Introduction

There exist in the literature many different definitions of image complexity. For example, it can be analyzed by using mathematical treatments based on Kolmogorov complexity theory [1]. Snodgrass et al. [2] refer to the visual complexity as the amount of detail or intricacy in an image. Birkhoff [3] relates the image complexity to visual aesthetics. Researchers from various fields have conducted psycho-physical experiments to study the subjective perception of visual complexity and some studies exist where experimental estimation of image complexity is correlated to objective measures. The state of the art studies differ in the

© Springer International Publishing Switzerland 2015
A. Trémeau et al. (Eds.): CCIW 2015, LNCS 9016, pp. 139–148, 2015.
DOI: 10.1007/978-3-319-15979-9_14

kind of stimuli used during the experimental sessions and on the type of objective measures used to correlate the subjective scores. Chikhman et al. [4] use Chinese hieroglyphs and outline images of well known common objects as stimuli sets. On the other side, experiments exist that address the image complexity of real world scenes, like the study by Oliva et al. [5]. Recently, Purchase et al. [6] use sixty images, including landscapes, domestic objects and city scenes, as stimuli. Further efforts attempt to describe the image complexity using different mathematical models like fuzzy approaches [7,8], information-theoretic approaches [9] and independent component analysis [10]. Rosenholtz et al. [11] associate the concept of complexity to that of visual clutter. They have tested three measures of visual clutter: Feature Congestion (FC), Subband Entropy (SE) and the edge density measure used to predict subjective judgments of image complexity by Mack and Oliva [12].

Image complexity can be useful in many different domains. It finds application to context-based image retrieval [10], icons and symbol search, particularly relevant in human computer interaction [13,14], and computer graphics, where a better understanding of visual complexity can aid in the development of more advanced rendering algorithms [15]. Other fields of application are image recognition [16], watermarking [17], compression [18], and image quality [19,20]. The image complexity concept is also used by neuroscientists, interested in the mechanisms of object recognition, learning and memory [21].

Aim of this paper is to investigate the role of color when evaluating image complexity. To this end, two types of experimental setups have been performed on a set of 29 color images and on the set of the corresponding grayscale images. These are real-world images, belonging to the image quality database LIVE [22]. Setup 1 is a ranking experiment, where observers rank the 29 images in increasing order of complexity. Setup 2 is a scaling experiment, where observers judge the image complexity on a continuous scale [0-100]. We point out that no definition of *image complexity* is provided to the observers during the experimental sessions. We investigate the effect of color on the perception of image complexity comparing the subjective results obtained with color and grayscale sets of images for both experimental setups. Moreover, we also consider ten image features as complexity measure candidates, and we evaluate their correlation with subjective data. The performance of these correlations are evaluated in terms of the linear Pearson Correlation Coefficient (PCC) and the Spearman Rank Order Correlation Coefficient (SROCC).

2 Color versus Grayscale: Subjective Data

The 29 images belonging to the LIVE database [22] have been used as stimuli for estimating subjective perception of image complexity. They have been chosen to sample different contents both in terms of low level features (frequencies, colors) and higher ones (face, buildings, close-up, outdoor, landscape). Their thumbnail color versions are shown in Figure1, and the corresponding greyscale versions are shown in Figure 2.

Fig. 1. Thumbnails of the color images used as stimuli in the psycho-physical experiments

Fig. 2. Thumbnails of the grayscale images used as stimuli in the psycho-physical experiments

2.1 Experimental Setup 1: Image Ranking

A group of 76 observers with normal or corrected-to-normal visual acuity and normal color vision took part in this psycho-physical experiment. Ishihara color test plates printed on paper have been preliminarily presented to the observers for detecting color vision deficiency. The images in the LIVE database were professionally printed on a high quality paper to create the cards for the psycho-physical experiment. The cards with the color images were given to 37 observers. Cards with the same greyscale images from the LIVE database were given to the remaining 39 observers. Observers could look at all the stimuli simultaneously for an unlimited time. The task of the observer was to arrange the images in order of increasing complexity. No definitions of complexity were imposed to the observers. The final rank of each image was obtained ranking the average of the positions assigned by the observers (from 1 the simplest image to 29 the most complex one).

2.2 Experimental Setup 2: Image Scaling

A group of 31 observers with normal or corrected-to-normal visual acuity and normal color vision took part in the psycho-physical experiment. A single stimulus method was adopted, where all the images are individually shown. No specific task was provided, just assessing the image complexity of each image using a scale in the range [0-100]. 14 observers evaluated the 29 color images, while the remaining 17 observers judged the grayscale counterparts. The images were shown on a web-based interface in a random order, different for each subject. The subjects reported their complexity judgments by dragging a slider onto a continuous scale. The position of the slider is automatically reset after each evaluation. A grayscale chart was shown to calibrate the brightness and the contrast of the monitor. Ishihara color test have been preliminarily presented to the observers for estimating color vision deficiency.

Seven training images were presented to the observers prior to the 29 test ones. These images have been used to train the subjects about the range of complexity to be evaluated. The corresponding data has been discarded and not considered as experimental result.

We have applied Z-score and outliers detection to obtain the final Mean Opinion Scores (MOS) of each image. The raw complexity score r_{ij} for the i-th subject ($i = 1, ...14$ in case of color images or $i = 1, ...17$ in case of grayscale images) and j-th image ($j = 1, ...29$) was converted into Z scores:

$$z_{ij} = \frac{r_{ij} - \bar{r_i}}{\sigma_i} \tag{1}$$

where $\bar{r_i}$ is the average of the complexity scores over all images ranked by the subject, and σ_i is the standard deviation. The Z scores were then averaged across subjects after the removal of the outlier scores. A score for an image was considered to be an outlier, and thus removed from the average computation, if it was outside an interval of width two standard deviations about the average score for that image.

3 Color versus Grayscale: Assessing Image Complexity

The following features have been considered as candidate complexity measures:

- **F1** Contrast, extracted applying the MATLAB function *graycoprops* to the gray-level co-occurence matrix.
- **F2** Homogeneity, Extracted applying the MATLAB function *graycoprops* to the gray-level co-occurence matrix.
- **F3** Edge density [12]: the MATLAB's Canny edge detector is applied to the image to measure the density of edge pixels.
- **F4** Feature Congestion [11]: its implementation involves: (1) computation of local measures (color, orientation, and luminance contrast) covariance at multiple scales and computing the volume of the local covariance ellipsoid, (2) combine clutter across scale and feature types, (3) pooling over space to get a single measure.

- **F5** Subband Entropy [11]: it is based on the notion of clutter as related to the efficiency with which the image can be encoded and inversely related to the amount of redundancy and grouping in the image.
- **F6** Compression Ratio of the image JPEG compressed with Q factor = 100.
- **F7** Number of Regions, calculated using the mean shift algorithm [23].
- **F8** Colorfulness [24]: linear combination of the mean and standard deviation of the pixel cloud in the color plane of CIELab.
- **F9** Number of colors [26]: number of distinct color in the image.
- **F10** Color harmony [25][26]: it is based on the perceived harmony of color combinations.

The first three features, labeled $F1$, $F2$, $F3$ work on grayscale images, features from $F4$ to $F7$ are mainly developed for color images but they are also meaningful for grayscale images while $F8$, $F9$, and $F10$ are meaningful only for color images.

These features are correlated with the subjective data obtained in the psychophysical experiments. In the case of the ranking experiment we are interested in assessing if the features are able to replicate the subjective ranks. In the case of the scaling experiment the aim is to assess the ability of the features to predict the MOS. For the latter case a proper logistic regression is used as follows.

Denoting by y_j the MOS of the $j-th$ image and by x_j the corresponding objective feature value, the logistic transformation reads:

$$f(x_j) = \frac{\alpha}{1 + e^{\beta(x_j - \gamma)}} + \delta \qquad (2)$$

where the parameters α, β, γ and δ are chosen to minimize the mean square error between the MOS $\{y_j\}$ and the predicted values $\{f(x_j)\}$.

4 Experimental Results

We initially investigate if color influences the perception of image complexity by analyzing the subjective data collected within each experimental setup. To this end we consider the raw data for both the experimental setups. In case of ranking experiment, we consider as raw data the average of the positions assigned by the observers. In case of scaling experiment, the raw data are the mean of the scores. The raw data collected for the grayscale images are correlated with those collected for the corresponding color images within each type of experiment. In Figure 3 on the left, the raw data of the ranking experiment are considered: the raw data of the 29 grayscale images are plotted with respect to the corresponding data of color images. In Figure 3 on the right, the raw data of the scaling experiment of the 29 grayscale images are plotted with respect to the raw data of the corresponding color versions. To measure the linear correlation between grayscale and color data for each experimental setup we evaluate the PCC, while to quantify their rank-order correlation we use the SROCC. In Table 1 these coefficients are reported for the two experimental setups.

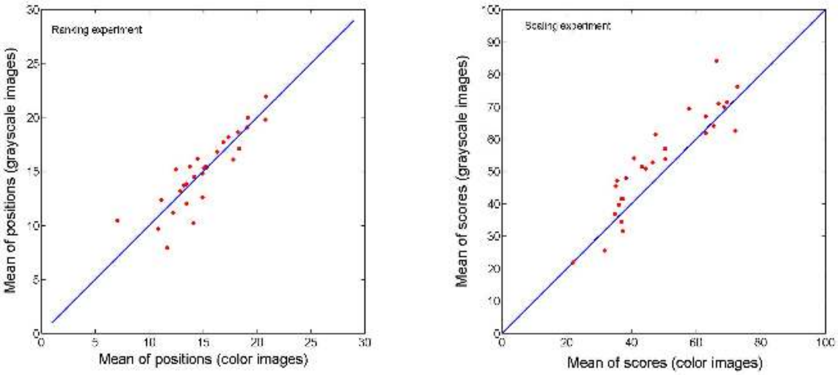

Fig. 3. Correlation between grayscale and color image data. Left: results from ranking experiment. Right: results from scaling experiment.

Table 1. Correlation coefficients between grayscale and color data for each experimental setup

Setup	PCC	SROCC
Ranking	0.877	0.903
Scaling	0.914	0.926

Subjective evaluations of color and grayscale images are highly correlated within each experimental setups. In particular in the case of scaling experiment these results suggest that the perception of image complexity is not significantly influenced by color. To have further insight into this issue, we evaluate the correlation between subjective and objective data. As objective data we adopted the ten features listed in Section 3. The last three features can not be evaluated for grayscale images as they are designed to measure color properties only.

In the case of the ranking experiment we consider the performance of the features in predicting the subjective rank. The results are presented using the SROCC and reported in Table 2 first row for the color images, second row for the grayscale images.

In the case of the scaling experiment we consider the performance of the features in predicting the MOS, using a proper logistic regression. The results are presented in terms of PCC and SROCC and reported in Tables 3 and 4.

In general we can notice that in the case of ranking experiment all the features evaluated on grayscale images better predict the subjective ranks than the corresponding ones on color images. Instead, in the case of scaling experiment all

Table 2. SROCC of the ten features in the ranking experiment

SROCC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Color	0.649	0.600	0.587	0.600	0.497	0.568	0.544	0.076	0.380	0.308
Grayscale	0.763	0.726	0.755	0.759	0.655	0.727	0.675	-	-	-

Table 3. PCC of the 10 features in the scaling experiment

PCC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Color	0.751	0.656	0.740	0.622	0.604	0.683	0.583	0.211	0.321	0.128
Grayscale	0.696	0.736	0.740	0.628	0.762	0.777	0.427	-	-	-

Table 4. SROCC of the 10 features in the scaling experiment

SROCC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Color	0.759	0.721	0.734	0.692	0.624	0.738	0.582	0.030	0.247	0.188
Grayscale	0.721	0.738	0.746	0.669	0.709	0.740	0.500	-	-	-

Fig. 4. Images in the first four rank positions (low complexity) in the ranking experiment for color and grayscale images

Fig. 5. Images in the last four rank positions (high complexity) in the ranking experiment for color and grayscale images

the features perform similarly for both color and grayscale data. This behavior is related to the higher correlation between color and grayscale data in the scaling experiment than in the ranking one (see Figure 3). The three color features $F8$, $F9$, and $F10$ are not appropriate to correlate subjective color data. This analysis suggests that the perception of image complexity is slightly influenced by color especially in the second experimental setup. The four lowest and four highest complexity images for both color and grayscale datasets and for both experimental setups are shown in Figures 4-7.

Fig. 6. Images with the four lowest MOS (low complexity) in the scaling experiment for color and grayscale images

Fig. 7. Images with the four highest MOS (high complexity) in the scaling experiment for color and grayscale images

5 Conclusions

In this work we have shown that there is a significant correlation between psycho-physical data on color and grayscale images when observers are asked to evaluate image complexity. This suggests that color does not influence significantly the perception of image complexity. Moreover, features that are developed only to measure color properties seem not to be suitable to correlate with the psycho-physical data. We recall that we have here considered real world images, where the lightness component provides enough information about the image content. Other kind of experiments, for example using images of color patches could yield different conclusions, as the grayscale images could be less meaningful. As a future work we plan to extend the psycho-physical experiments both in number of observers and in number of images. However psycho-physical experiments with a huge amount of images are a difficult task. In fact images should be divided into different groups to be judged by different groups of observers and the final data should be properly aligned. Furthermore we plan to investigate if a proper combination of metrics that takes into account simultaneously spatial, frequency and color image characteristics can better predict subjective evaluations.

References

1. Kolmogorov, A.N.: Three approaches to the quantitative definition of information. Problems of Information Transmission **1**(1), 1–7 (1965)
2. Snodgrass, J.G., Vanderwart, M.: A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory **6**(2), 174 (1980)
3. Birkhoff, G.D.: Collected mathematical papers (1950)
4. Chikhman, V., Bondarko, V., Danilova, M., Goluzina, A., Shelepin, Y.: Complexity of images: Experimental and computational estimates compared. Perception **41**, 631–647 (2012)
5. Oliva, A., Mack, M.L., Shrestha, M.: Identifying the perceptual dimensions of visual complexity of scenes. In: Proc. 26th Annual Meeting of the Cognitive Science Society (2004)
6. Purchase, H.C., Freeman, E., Hamer, J.: Predicting visual complexity. In: Proceedings of the 3rd International Conference on Appearance, Edinburgh, UK, pp. 62–65 (2012)
7. Mario, I., Chacon, M., Alma, D., Corral, S.: Image complexity measure: a human criterion free approach. In: Annual Meeting of the North American Fuzzy Information Processing Society, NAFIPS 2005, pp 241–246. IEEE (2005)
8. Cardaci, M., Di Gesú, V., Petrou, M., Tabacchi, M.E.: On the evaluation of images complexity: a fuzzy approach. In: Bloch, I., Petrosino, A., Tettamanzi, A.G.B. (eds.) WILF 2005. LNCS (LNAI), vol. 3849, pp. 305–311. Springer, Heidelberg (2006)
9. Rigau, J., Feixas, M., Sbert, M.: An information-theoretic framework for image complexity. In: Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 177–184. Eurographics Association (2005)
10. Perkiö, J., Hyvärinen, A.: Modelling image complexity by independent component analysis, with application to content-based image retrieval. In: Alippi, C., Polycarpou, M., Panayiotou, C., Ellinas, G. (eds.) ICANN 2009, Part II. LNCS, vol. 5769, pp. 704–714. Springer, Heidelberg (2009)
11. Rosenholtz, R., Li, Y., Nakano, L.: Measuring visual clutter. Journal of Vision **7**(2), 17 (2007)
12. Mack, M.L., Oliva, A.: Computational estimation of visual complexity. In: The 12th Annual Object, Perception, Attention, and Memory Conference (2004)
13. Reppa, I., Playfoot, D., McDougall, S.J.P.: Visual aesthetic appeal speeds processing of complex but not simple icons. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 52, pp. 1155–1159. SAGE Publications (2008)
14. Forsythe, A.: Visual complexity: is that all there is? In: Harris, D. (ed.) EPCE 2009. LNCS, vol. 5639, pp. 158–166. Springer, Heidelberg (2009)
15. Ramanarayanan, G., Bala, K., Ferwerda, J.A., Walter, B.: Dimensionality of visual complexity in computer graphics scenes. In: Electronic Imaging 2008, pp. 68060E–68060E. International Society for Optics and Photonics (2008)
16. Peters, R.A., Strickland, R.N.: Image complexity metrics for automatic target recognizers. In: Automatic Target Recognizer System and Technology Conference, pp. 1–17 (1990)
17. Yaghmaee, F., Jamzad, M.: Estimating watermarking capacity in gray scale images based on image complexity. EURASIP Journal on Advances in Signal Processing **2010**, 8 (2010)

18. Yu, H., Winkler, S.: Image complexity and spatial information. In: 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX), pp. 12–17. IEEE (2013)
19. Corchs, S., Gasparini, F., Schettini, R.: Grouping strategies to improve the correlation between subjective and objective image quality data. In: Image Quality and System Performance X, IS&T/SPIE Electronic Imaging, p. 86530D(1–8). SPIE (2013)
20. Bianco, S., Ciocca, G., Marini, F., Schettini, R.: Image quality assessment by preprocessing and full reference model combination. In: Image Quality and System Performance VI, vol. 7242, p. 72420O. SPIE (2009)
21. Donderi, D.C.: Psychological Bulletin. Visual complexity: a review **132**(1), 73 (2006)
22. Sheik, H., Wang, Z., Cormakc, L., Bovik, A.: LIVE Image Quality Assessment Database Release 2. http://live.ece.utexas.edu/research/quality
23. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis and the edge detection algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**, 603–619 (2002)
24. Hasler, D., Suesstrunk, S.E.: Measuring colorfulness in natural images. Electronic Imaging **2003**, 87–95 (2003)
25. Solli, M., Lenz, R.: Color harmony for image indexing. In: IEEE 12th International Conference on Computer Vision Workshops, pp. 1885–1892 (2009)
26. Artese, M.T., Ciocca, G., Gagliardi, I.: Good 50x70 Project: A portal for Cultural And Social Campaigns. In: IS&T Archiving 2014 Conference, Final Program and Proceedings, pp. 213–218 (2014)