Acta Psychologica 160 (2015) 43–57

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Computerized measures of visual complexity☆

Penousal Machado [a,*], Juan Romero [b], Marcos Nadal [c], Antonino Santos [b], João Correia [a], Adrián Carballal [b]

[a] CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
[b] Faculty of Computer Science, University of A Coruña, Spain
[c] Department of Basic Psychological Research and Research Methods, University of Vienna, Austria

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form 21 May 2015
Accepted 18 June 2015
Available online 10 July 2015

Keywords:
Visual complexity
Psychological aesthetics
Vision
Machine learning

ABSTRACT

Visual complexity influences people's perception of, preference for, and behaviour toward many classes of objects, from artworks to web pages. The ability to predict people's impression of the complexity of different kinds of visual stimuli holds, therefore, great potential for many domains, basic and applied. Here we use edge detection operations and several image metrics based on image compression error and Zipf's law to estimate the visual complexity of images. The experiments involved 800 images, each previously rated by thirty participants on perceived complexity. In a first set of experiments we analysed the correlation of individual features with the average human response, obtaining correlations up to $r_s = .771$. In a second set of experiments we employed Machine Learning techniques to predict the average visual complexity score attributed by humans to each stimuli. The best configurations obtained a correlation of $r_s = .832$. The average prediction error of the Machine Learning system over the set of all stimuli was .096 in a normalized 0 to 1 interval, showing that it is possible to predict, with high accuracy human responses. Overall, edge density and compression error were the strongest predictors of human complexity ratings.

1. Introduction

People's preferences for visual objects, scenes, and displays are the result of various cognitive and affective processes (Chatterjee, 2004; Leder, Belke, Oeberst, & Augustin, 2004). Research has shown that several perceptual features—such as colour, colour combinations, contour, or symmetry—influence people's visual preferences and affective responses (Bertamini, Palumbo, Gheorghes, & Galatsidas, in press; Palmer, Schloss, & Sammartino, 2013; Pecchinenda, Bertamini, Makin, & Ruta, 2014). One of such features, complexity, is believed to have a strong impact on preference and affect, given its relation to arousal (Berlyne, 1971; Marin & Leder, 2013), and has therefore been awarded central roles in psychological models of aesthetic appreciation (Berlyne, 1971; Fechner, 1876). From a basic science perspective, thus, research on how the perceptual features that contribute to visual complexity are processed, and how this processing leads to liking and other affective responses, increases our understanding of one of our species distinctive traits: the capacity for aesthetic appreciation. From an applied perspective, it has implications for the design of architectural spaces (Heath, Smith, & Lim, 2000; Imamoglu, 2000), advertisements (Pieters, Wedel, & Batra, 2010), packages (Reimann, Zaichkowsky, Neuhaus, Bender, & Weber, 2010), web pages (Bauerly & Liu, 2008; Krishen, Kamra, & Mac, 2008; Lavie & Tractinsky, 2004; Moshagen & Thielsch, 2010), and in-vehicle navigation devices (Lavie, Oron-Gilad, & Meyer, 2011), among other domains, where visual complexity impacts both liking and usability.

It has long been believed that two aspects of complexity, order and variety, determine beauty. From this perspective, beauty emerges from "unity in variety" (Tatarkiewicz, 1972). The importance of two different—and sometimes opposing—forces was introduced into experimental psychology by Fechner (1876), who formulated the "principle of unitary connection of the manifold" (Cupchik, 1986), that argued that stimuli are pleasing when they adequately balance complexity and order. Birkhoff (1932) formulated this relation between order and complexity in mathematical terms, and argued that beauty increased with order and decreased with complexity. He defined order on the basis of repetition and redundancy, and complexity as an expression of numerousness. Eysenck's (1941, 1942) studies on the correlation between the aesthetic measure predicted by Birkhoff's (1932) formula and participants' beauty ratings suggested that both order and complexity contribute positively to the appreciation of beauty.

Berlyne (1970, 1971) was probably the first to provide a proper psychobiological explanation for the effects of complexity on preference. Berlyne (1971) posited that the hedonic state resulting from the interaction of reward and aversion brain systems would lead people to prefer intermediate levels of complexity, which was defined according to such

☆ Penousal Machado was supported by the Portuguese Foundation for Science and Technology in the scope of project SBIRC (PTDC/EIA–EIA/115667/2009). Marcos Nadal was supported by research grant Research Grant FFI2010-20759 awarded by the Spanish Ministerio de Ciencia e Innovación.

* Corresponding author at: Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal.

E-mail address: machado@dei.uc.pt (P. Machado).

http://dx.doi.org/10.1016/j.actpsy.2015.06.005
0001-6918/

aspects as pattern regularity, amount of elements, their heterogeneity, or the irregularity of the forms (Berlyne, 1963, 1970, 1971; Berlyne, Ogilvie, & Parham, 1968). In Berlyne's framework, order is not orthogonal to complexity, given that disorganization is regarded as a kind of complexity, together with the amount of elements. Several studies were conducted to test this hypothesis, employing diverse visual stimuli. Recent research has shown that their results were strongly conditioned by the way complexity had been defined, manipulated and measured (Nadal, Munar, Marty, & Cela-Conde, 2010).

2. Measuring complexity

It has been known for some time now that people's perception of complexity is not merely a direct reflection of the complexity inherent to visual stimuli. Attneave (1957) noted that “the amount of information contained in a stimulus (from the experimenter's point of view) may vary greatly without changing the apparent complexity of the stimulus” (Attneave, 1957, p. 225). Perception is a constructive process. Although it is based on sensory information, its purpose is not to render the world as it is, but to provide us with an image that we can understand and is coherent with our prior knowledge about the world. In order to do so, perception is guided by inference, hypotheses, and other top-down processes, as well as context, which can strongly influence the appearance of an object. Gestalt psychologists characterized several perceptual processes whereby visual features are joined, segregated and grouped to construct meaningful images, and these processes have a crucial role in determining perceived complexity (Strother & Kubovy, 2003).

Even Berlyne (1974) emphasized “The collative variables [including complexity] are actually subjective, in the sense that they depend on the relations between physical and statistical properties of stimulus objects and processes within the organism. A pattern can be more novel, complex, or ambiguous for one person than for another or, for the same person, at one time than at another”. “Nevertheless — he added — many experiments, using rating scales and other techniques, have confirmed that collative properties and subjective informational variables tend, as one would expect, to vary concomitantly with the corresponding objective measures of classical information theory” (Berlyne, 1974, p19).

In principle, thus, it should be possible to arrive at a computational measure of visual complexity. This constitutes an interesting objective for at least two reasons. In a basic sense, measures of images' intrinsic complexity would enable determining the perceptual, cognitive or contextual features that influence perceived complexity, moving closer or away from the objective (computational) measure. In an applied sense, it would allow researchers, designers and engineers to anticipate participants', consumers' and users' aesthetic and affective responses to the complexity in their products, ranging from web pages to architectural facades, and including visual displays of all sorts. This would greatly save the time and costs related with post-production tests and surveys.

One of the most popular way of determining visual complexity has been to derive a set of normative scores by asking large samples of participants to rate sets of stimuli on a number of scales, including complexity (Alario & Ferrand, 1999; Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Snodgrass, 1997). This method, however, has a number of drawbacks. First, people's rating of complexity can be confounded by familiarity (Forsythe, Mulhern, & Sawey, 2008) and style (Nadal et al., 2010). Second, it is only useful for images that have already been produced, and does not allow the prediction of the perceived complexity of images whose production is being planned or under development. In this sense, algorithms represent more fruitful and practical avenue possibility.

In their study on icon abstractness García, Badre, and Stasko (1994) developed an algorithmic measure of complexity. This measure took into account the amount of horizontal, vertical, and diagonal lines, as well as the number of open and closed figures, and letters in each icon. McDougall, Curry, and de Bruijn (1999) used the same measure to quantify the complexity of a new set of figures, and they showed that it correlated well with people's judgement of visual complexity (McDougall, de Bruijn, & Curry, 2000). Given how time consuming it was to calculate this metric, Forsythe, Sheehy, and Sawey (2003) devised an automated system to measure icon complexity. They based this metric on perimeter detection measures and a structural variability measure. Their results showed strong correlations between their metric and the scores provided by García et al. (1994) and McDougall et al.'s (1999) studies, revealing that it is possible to approximate human appraisals of complexity with computational metrics of structural properties of images.

The main drawback of this kind of metrics is its limited application to relatively simple and isolated icons and symbols. Algorithmic measures of complexity for richer stimuli, like pictures from nature, chart displays and art, have tended to be based on algorithmic information theory (Donderi, 2006). In short, this theory postulates that the minimum length of the code required to describe a visual image constitutes a good measure of its complexity (Leeuwenberg, 1969; Simon, 1972). Donderi (2003) showed that compressed file size was a good approximation to this minimum length. Furthermore, JPEG and ZIP compressed file lengths significantly correlated with subjectively rated complexity and predicted search time and errors in tasks involving chart displays (Donderi & McFadden, 2005).

Computational measures have also been applied to attempt to quantify the complexity of artworks. Forsythe, Nadal, Sheehy, Cela-Conde, and Sawey (2011) examined the correlation between people's judgement of complexity for 800 artistic and nonartistic, abstract and representational, visual stimuli and JPEG and GIF compression measures, as well as with a perimeter detection measure. Their results showed that the three computational measures significantly correlated with judged complexity, with GIF compression exhibiting the strongest relation ($r_s = .74$) and perimeter detection the weakest ($r_s = .58$), though there were certain differences according to the kind of stimuli.

Marin and Leder (2013) also compared the extent to which several computational measures correlated with participants' complexity ratings of different kinds of materials. For a subset of stimuli from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2005), they found that TIFF file size ($r_s = .53$) and JPEG file size ($r_s = .52$) correlated strongest with subjective complexity ratings. Similarly to Forsythe et al.'s (2011) work, Marin and Leder (2013) reported that measures of perimeter detection showed weaker correlations ($r_s \sim .44$). For this set of stimuli, the highest correlations were obtained with an edge detection measure: the root mean square contrast (RMS), related to the presence of high-contrast features. In this case, the correlation between complexity ratings and the images' mean contrast values of the RMS contrast map was $r_s = .59$. Interestingly, these results were not mirrored in Marin and Leder's (2013) second experiment, which aimed to examine the relation between the same measures and human complexity ratings of 96 representational paintings. For this set, none of the compressed file size measures correlated significantly with the ratings. In fact, the only measure to correlate significantly with complexity ratings was the standard deviation of the mean values of edge detection based on phase congruency ($r_s \sim .38$).

The discrepancies between Forsythe et al.'s (2011) and Marin and Leder's (2013) results probably have to do with the selected materials and procedure. Whereas Forsythe et al. (2011) excluded affectively moving images, Marin and Leder (2013) selected the images in the two aforementioned experiments to accomplish a balanced variation along the arousal and pleasantness dimensions. The images used by Forsythe et al. (2011) were selected on the basis of pilot experiments to cover a broad range of visual complexity, understood in a general sense as the degree of intricacy; the ones used by Marin and Leder (2013) were either figure-ground compositions or complex visual scenes.