

RESEARCH ARTICLE

Predicting Complexity Perception of Real World Images

Silvia Elena Corchs^{1,3*}, Gianluigi Ciocca^{1,3}, Emanuela Bricolo^{2,3}, Francesca Gasparini^{1,3}

1 Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy, **2** Department of Psychology, University of Milano-Bicocca, Via dell' Innovazione 10, 20126 Milano, Italy, **3** Milan Center for Neuroscience, Milano, Italy

* silvia.corchs@disco.unimib.it

Abstract

The aim of this work is to predict the complexity perception of real world images. We propose a new complexity measure where different image features, based on spatial, frequency and color properties are linearly combined. In order to find the optimal set of weighting coefficients we have applied a Particle Swarm Optimization. The optimal linear combination is the one that best fits the subjective data obtained in an experiment where observers evaluate the complexity of real world scenes on a web-based interface. To test the proposed complexity measure we have performed a second experiment on a different database of real world scenes, where the linear combination previously obtained is correlated with the new subjective data. Our complexity measure outperforms not only each single visual feature but also two visual clutter measures frequently used in the literature to predict image complexity. To analyze the usefulness of our proposal, we have also considered two different sets of stimuli composed of real texture images. Tuning the parameters of our measure for this kind of stimuli, we have obtained a linear combination that still outperforms the single measures. In conclusion our measure, properly tuned, can predict complexity perception of different kind of images.



OPEN ACCESS

Citation: Corchs SE, Ciocca G, Bricolo E, Gasparini F (2016) Predicting Complexity Perception of Real World Images. PLoS ONE 11(6): e0157986. doi:10.1371/journal.pone.0157986

Editor: Feng Shao, Ningbo University, CHINA

Received: December 10, 2015

Accepted: May 9, 2016

Published: June 23, 2016

Copyright: © 2016 Corchs et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The study of image complexity perception can be useful in many different domains. Within the human-computer interaction field, Forsythe et al. [1] proposed an automated system to predict perceived complexity and applied it in icon design and usability. Reinecke et al. [2] quantified visual complexity of website screenshots, formulating a model for the prediction of visual appeal in order to improve the user experience on the web. In addition it has been deemed useful in computer graphics, where a better understanding of visual complexity can aid the development of more advanced rendering algorithms [3] or image based 3D reconstruction [4]. Digital watermarking methods also can benefit from an estimation of image complexity as it has been related to the amount of information that can be hidden in image [5]. Nowadays it finds application in context-based image retrieval [6] where a model of joint complexity of images gives distances that can be used to estimate the degree of similarity between images. For instance in the research

area of mobile visual search an hypothetical image complexity coefficient could aid within the image search process [7–9]. Within video quality research, the compression level and bandwidth allocation can be related to image complexity: low complexity stimuli can be compressed more easily and requires less bandwidth than an image with high complexity [10]. In the field of Automatic Target Recognizer, Peters and Strickland [11] proposed an image complexity metric to provide an a priori estimate of the difficulty of locating a target in an image. Moreover, object complexity is also considered as an object-level feature to predict saliency in the visual field [12, 13]. The authors introduced different attributes to predict saliency, and among them complexity. The image complexity concept is also used by neuroscientists interested in the mechanisms of recognition, learning and memory [14]. Recently, the complexity topic has been investigated across domains, using both visual and musical stimuli [15].

The different definitions of image complexity found in the literature depend on the specific task and the application domain. From a purely mathematical point of view, Kolmogorov [16] defined the complexity of an object as the length of the shortest program that can construct the object from basic elements, or description language. Snodgrass et al. [17] referred to the visual complexity as the amount of detail or intricacy in an image. Heaps and Hande [18] defined complexity as the degree of difficulty in providing a verbal description of an image. Image complexity is also related to aesthetics: Birkhoff [19] introduced the concept of aesthetic measure as the ratio between order and complexity.

As reviewed by [20], image complexity should be considered in relation to other factors like familiarity, novelty, and interest among others. Complexity is influenced not only by the spatial properties of the stimuli but also by the temporal dimension [21, 22]. Attempts to describe the image complexity using different mathematical models have been proposed, like for example fuzzy approaches [21, 23–25], independent component analysis [6] and information theory-based approaches [26].

Visual complexity can be also related to the concept of visual clutter. Rosenholtz et al. [27] proposed two measures of visual clutter: Feature Congestion (FC) and Subband Entropy (SE). They compared FC and SE performances with respect to the edge density used by Mack and Oliva [28].

Several researchers have conducted experiments to study the subjective perception of visual complexity. The state of the art studies differ in the kind of stimuli used during the experimental sessions and on the type of objective measures used to correlate the subjective scores.

During the experiments by Chikhman et al. [29], observers were asked to rank the complexity of two groups of stimuli: unfamiliar Chinese hieroglyphs and outline images of well-known common objects. To predict image complexity, they considered: spatial characteristics of the images, spatial-frequency characteristics, a combination of spatial and Fourier properties, and the size of the image encoded as a JPEG file. Oliva et al. [30] analyzed image complexity of indoor scenes. During their experimental session, participants performed a hierarchical grouping task in which they divide scenes into successive groups of decreasing complexity, describing the criteria they used at each stage. Their results demonstrated a multi-dimensional representation of visual complexity (quantity of objects, clutter, openness, symmetry, organization, variety of colors). Purchase et al. [31] conducted an experimental study using landscapes, domestic objects and city scenes as stimuli. The authors correlated their experimental results to a set of four computational metrics: colour, edges of objects, intensity variation, and file size. Cavalcante et al. [32] proposed a method to evaluate the complexity perceived in streetscapes images based on the statistics of local contrast and spatial frequency. Ciocca et al. [33] investigated the influence of color in the perception of image complexity, considering several visual features, that measure colors as well as other spatial properties. The perception of texture complexity has been studied by different authors [34–36]. Recently, Guo et al. [37] assessed the visual

complexity of paintings, providing a machine learning scheme for investigating the relationship between human visual complexity perception and low-level image features.

Taking into account the multi-dimensional aspect of complexity, we propose a complexity measure based on a combination of several features related to spatial, frequency and color properties in order to predict complexity perception of real world images. We here consider two different kind of real world stimuli: Real Scenes, (RS) and real texture patches (TXT). The aim of our work is to propose a general purpose metric, that, tuned with respect to the kind of stimuli considered, can better correlate the subjective data with respect to single measures. We here propose a linear combination of visual features to predict image complexity perception where the weighting coefficients can reveal the role of each of them. Starting from a given set of stimuli, we apply a Particle Swarm Optimization (PSO) [38, 39] to find the weighting coefficients of the linear combination that best fits subjective data. We set up an experiment, where observers evaluated the complexity of real world scenes on a web-based interface and they were also asked to verbally describe the criteria that guided their evaluation.

Analyzing the most common criteria reported by the observers in the questionnaire, it could be possible to associate some of them with single image features and compare the frequency of these criteria to the weighting coefficients of the linear combination. To test our proposal we performed a second experiment on a new database of real world scenes, where the linear combination previously obtained is correlated with new subjective data. To verify the usefulness of our complexity measure to predict complexity of a different type of stimuli, we performed two more experiments on two different datasets of texture images. We have chosen texture images because they present a high range of complexity levels like real world scenes but with a different semantic content. We again apply PSO to find the new weighting coefficients of the linear combination proposed on the first of these two texture sets of images and we test the obtained measure on the other texture dataset. Up to our knowledge no supervised or unsupervised measure to evaluate complexity perception of real world images has been presented in the literature. Thus as a benchmark for evaluating our proposal, we consider two measures of visual clutter: FC and SE [27]. We have chosen these two measures as they have been frequently used in the literature, where they have shown correlation with image complexity perception, [3, 32, 40], even if they are not unsupervised measures and they were not specifically designed to predict image complexity.

Materials and Methods

In this work we performed four experiments where the task is to evaluate image complexity. Each experiment is characterized by a different set of visual stimuli.

Stimuli

The images used as stimuli are all of high quality, acquired with professional and semi-professional cameras.

In Experiment 1 we used 49 images depicting real world scenes (RS1 dataset) belonging to the personal photo collection of the authors (RSIVL [41]). For Experiment 2 we considered other 49 real world scenes (RS2 dataset). These images correspond to the reference high quality images of the LIVE [42–44] (29 images) and the IVL database [45, 46] (20 images). Images belonging to RS1 and RS2 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). They include pictures of faces, people, animals, closeup shots, wide-angle shots, nature scenes, man-made objects, images with distinct foreground/background configurations, and images without any specific object of interest. Experiments 3 and 4 consider two different

datasets of real texture images, that represent a kind of stimuli with contents significantly different from those represented within RS1 and RS2.

In Experiment 3, we consider 54 real texture images (TXT1 dataset), belonging to the Vis-Tex data set [47]. This data set consists of 864 images representing 54 classes of natural objects or scenes captured under non-controlled conditions with a variety of devices. From each of the 54 classes, one image has been chosen as representative of the corresponding group. In Experiment 4 we use texture images belonging to the Raw Food Texture database (RawFooT) [48, 49]. It includes images of 68 samples of food textures, acquired under 46 lighting conditions. In our work we have used as stimuli 58 texture images acquired under the D65 lighting condition and frontal direction (TXT2 dataset).

Participants

Participants were recruited from the Informatics Department of the University of Milano Bicocca and were either students, researchers or administrative employees. No participants under the age of 18 were involved in our study and no health or medical data was collected from participants.

Through the web interface, informed consent was given by all participants. The data was collected anonymously.

For all experiments, six Ishihara tables were preliminarily presented to the observers to estimate color vision deficiency. If the participants did not report correctly any of the six they were discarded from the subjects' pool.

In Experiment 1, 28 observers participated, and 2 of them failed the Ishihara test. 26 observers remained: 15 women, mean age 37, range 18–53. In Experiment 2, 39 observers different from those of Experiment 1 were involved, and 3 of them failed the Ishihara test. 36 observers remained: 16 women, mean age 33, range 18–53. In Experiment 3, from the 19 initial observers, 17 remained after the control test: 5 women, mean age 35, range 23–51. In Experiment 4, from the 25 initial observers, 23 remained after the control test: 10 women, mean age 28, range 19–68.

All the experiments reported in this article were conducted in accordance with the Declaration of Helsinki and the local guidelines of the University of Milano Bicocca (Italy). No ethical approval was required for the present study. All the stimuli and subjective data are available at our web site [50].

Experimental setup

In all four experiments observers were asked to judge images individually presented on a web-interface.

Before the start of the experiment, a grayscale chart was shown to allow the observers to calibrate the brightness and the contrast of the monitor. The observers were asked to regulate the contrast of their monitor to distinguish the maximum number of bands and discern details in shadows and in highlights. In Fig 1 we report the web-interface and the contrast chart used in the experiment.

After calibration, the stimuli were shown in random order, different for each subject. Subjects reported their complexity judgment (score) by dragging a slider onto a continuous scale in the range [0–100]. Stimuli were presented for an unlimited time, up to response submission. The position of the slider was automatically reset after each evaluation at the midpoint of the scale.

In order to get the observers accustomed to the experiment, seven practice trials were presented at the beginning of each experiment, with images not included in the dataset. The corresponding data were discarded and not considered for any further analysis. At the end of the experimental session, the observers were asked to verbally describe the characteristics of the stimuli that affected their evaluation of visual complexity.

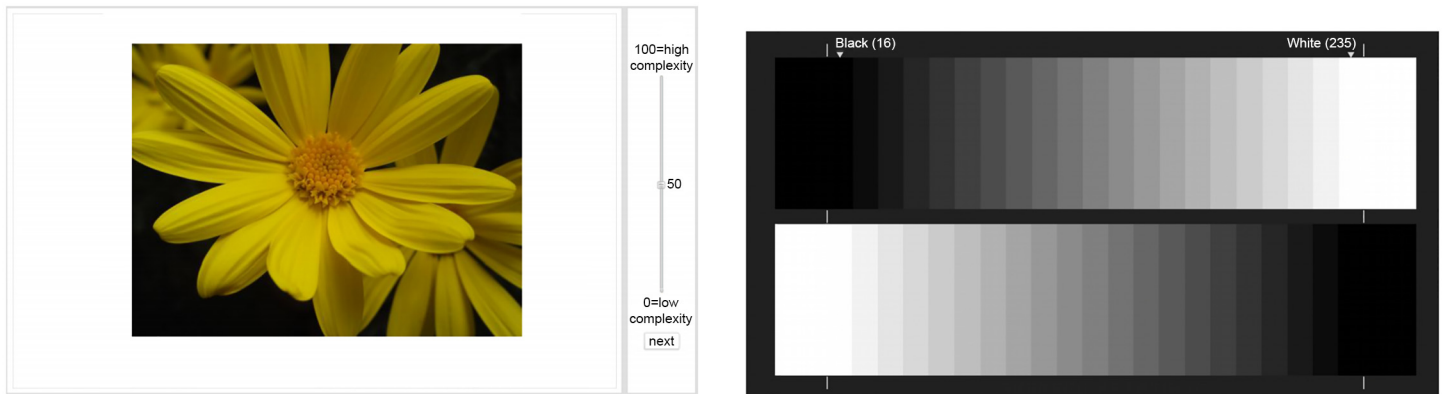


Fig 1. Web-interface of the experiment (left) and contrast chart used to calibrate the monitor (right).

doi:10.1371/journal.pone.0157986.g001

Subject scoring

Mean subjective scores were computed for each observer. The raw, subjective, complexity score r_{ij} for the i -th subject ($i = 1, \dots, S$, with S = number of subjects) and j -th image I_j ($j = 1, \dots, N$, with N = number of dataset images) was converted into its corresponding Z-score as follows:

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

where \bar{r}_i and σ_i are the mean and the standard deviation of the complexity scores over all images ranked by the i -th subject.

Data were cleaned using a simple outlier detection algorithm. A score for an image was considered to be an outlier if it fell outside an interval of two standard deviations width about the mean score for that image across all subjects.

The remaining Z-scores, were then averaged across subjects to yield the mean scores y_j for each image j :

$$y_j = \frac{1}{S} \sum_{i=1}^S z_{ij} \tag{2}$$

Our proposal of complexity measure

Due to the multi-dimensional aspect of complexity, we here propose a complexity measure based on a linear combination of K different features related to spatial, frequency and color properties. This Linear Combination (LC) can be written as follows:

$$LC(I_j) = \sum_{k=1}^K a_k \times \mathcal{M}_k(I_j) \tag{3}$$

where I_j is the j -th image of the considered dataset ($j = 1, \dots, N$), and \mathcal{M}_k is the measure of the k -th feature. Let $x_j = LC(I_j)$ and $x_{kj} = \mathcal{M}_k(I_j)$, Eq (3) can be rewritten in a compact way as:

$$x_j = \sum_{k=1}^K a_k \times x_{kj} \tag{4}$$

The set of optimal parameters $\{a_k\} = A^* \in \mathbb{R}^K$ of Eq (4) were chosen in order to optimally fit subjective data using a population based stochastic optimization technique, called Particle Swarm Optimization (PSO) [38, 39].

In PSO, a population of individuals is initialized as random guesses to the problem solution and a communication structure is also defined, assigning neighbors for each individual to interact with. These individuals are candidate solutions. An iterative process to improve these candidate solutions is set in motion. The particles iteratively evaluate the fitness of the candidate solutions and remember the location where they had their best success. The individual's best solution is called the particle best. Each particle makes this information available to its neighbors. They are also able to see where their neighbors have had success. Movements through the search space are guided by these successes. The swarm is typically modeled by particles in multidimensional space that have a position and a velocity. These particles fly through hyperspace and have two essential reasoning capabilities: their memory of their own best position and their knowledge of the global or their neighborhood's best position. Members of a swarm communicate good positions to each other and adjust their own position and velocity based on these good positions.

Recalling that one of the criteria widely used to evaluate the performance of a measure to fit subjective data is the linear Pearson Correlation Coefficient (PCC), we have chosen it as the fitness function to be maximized. To take into account the non linear mapping between objective and subjective data, the complexity measure x_j is previously transformed using a logistic function f [51].

The fitness function is thus:

$$PCC(A) = \frac{\sum_{j=1}^N (f(x_j) - \overline{f(x)}) (y_j - \bar{y})}{\sqrt{\sum_{j=1}^N (f(x_j) - \overline{f(x)})^2} \sqrt{\sum_{j=1}^N (y_j - \bar{y})^2}} \tag{5}$$

where A is a feasible solution, $f(x_j)$ is the logistically transformed value of the combined objective measure LC for the j -th image, and $\overline{f(x)}$ and \bar{y} are the means of the respective data sets.

The optimal parameter values A^* are thus obtained as:

$$A^* = \max_{A \in \mathbb{R}^K} (r(A)) \tag{6}$$

Note that our fitness function introduces a simple form of regularization of the searched model that is able to mitigate possible overfitting. In fact, optimizing the PCC defined in Eq (5) we are looking for the solution that minimized the square errors between subjective data and a monotonic curve described by the logistic function.

To benchmark our proposal we consider two clutter measures developed by Rosenholtz et al. [27]. Both of them are defined as a combination of different image features. They were not specifically designed to predict image complexity but have been frequently used in the literature as they show good correlation with complexity perception [3, 32]. The MATLAB implementation provided by the authors has been used:

- Feature Congestion (FC): three clutter maps for the image, representing color, texture and orientation congestion are evaluated across scales and properly combined to get a single measure.

- Subband Entropy (*SE*): it is related to the number of bits required for subband image coding. After decomposing the luminance and the chrominance channels into wavelet subbands, the entropy is computed within each band and a weighted sum of these entropies is proposed as clutter measure.

Objective measures

In what follows we list and briefly describe the 11 measures used in the linear combination of Eq (3). These measures evaluate simple visual features and were chosen as they were already used as complexity measures in the literature or could be related to complexity. The first six are computed on grayscale images and do not take into account the color information. \mathcal{M}_1 to \mathcal{M}_4 measure properties of the Grey Level Co-occurrence Matrix (GLCM), which is one of the earliest techniques used for image texture analysis and classification [52] and the MATLAB function *graycoprops* was used to compute them.

\mathcal{M}_1 : Contrast; it measures the intensity contrast between a pixel and its neighbors over the whole image.

\mathcal{M}_2 : Correlation; it measures how correlated a pixel is to its neighbors over the whole image.

\mathcal{M}_3 : Energy; it is the sum of squared elements in the GLCM.

\mathcal{M}_4 : Homogeneity, it measures the closeness of the distribution of elements in the GLCM with respect to the GLCM diagonal.

\mathcal{M}_5 : Frequency Factor, it is the ratio between the frequency corresponding to the 99% of the image energy and the Nyquist frequency [53].

\mathcal{M}_6 : Edge Density, it is obtained applying the Canny edge detector to the grayscale image with the parameters indicated by Rosenholtz et al. [27].

The following two \mathcal{M}_7 to \mathcal{M}_8 , describe image features which take into account color information when present:

\mathcal{M}_7 : Compression Ratio, it is here evaluated as the ratio of the image JPEG compressed with Q factor = 100 and the full size uncompressed image [54].

\mathcal{M}_8 : Number of Regions calculated using the mean shift algorithm [55].

Finally, measures from \mathcal{M}_9 to \mathcal{M}_{11} evaluate mainly color image properties:

\mathcal{M}_9 : Colorfulness; it consists in a linear combination of the mean and standard deviation of the pixel cloud in the color plane [56].

\mathcal{M}_{10} : Number of Colors; measures the number of distinct color in the RGB image, as described in [57].

\mathcal{M}_{11} : Color Harmony; based on the perceived harmony of color combinations. It is composed of three parts: the chromatic effect, the luminance effect, and the hue effect [58] [57].

Results

Experiment 1

In Experiment 1 we use the 49 images of real world scenes, belonging to the RS1 dataset. The subjective data collected is processed to obtain the mean scores (see Eq (2)). The RS1 images,

Table 1. Parameters for Eq (3), obtained averaging the A^* parameters over 1000 runs of PSO.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
0.2076	0.1986	-0.4046	0.1305	0.5548	-0.1245	-0.1257	0.6167	-0.2588	0.4645	-0.0692

doi:10.1371/journal.pone.0157986.t001

Since the single measures used to obtain the linear combination LC_{RS1} have been previously normalized, from Table 1 we can infer the role of each of them when predicting image complexity. The highest contribution to the linear combination comes from \mathcal{M}_8 (Number of regions), followed by \mathcal{M}_5 (Frequency Factor) and \mathcal{M}_{10} (Number of Colors) while measure \mathcal{M}_{11} (Harmony) is the one with lowest weight. The sign of the coefficients mainly depends on two different aspects. The first one is related to how each single measure correlates with the subjective evaluations. In Fig 3 the scatter plots between mean scores and each of the 11 single measures are shown together with the monotonic functions that best fit the data. Some of the measures show a monotonically increasing correlation while other a monotonically decreasing one. The second aspect is related to the partial correlation between some features. A minus in the linear combination can also take into account the attempt of the PSO algorithm to reduce the redundancy.

In Fig 4 the scatter plot between subjective (mean scores) and objective (LC_{RS1}) data of the RS1 images is reported. The monotone function that best fits the data is also shown with a continuous line. To benchmark our proposal, we also plot in the same Figure the scatter plots between mean scores and FC and SE respectively. To quantify the performance of these complexity measures to correlate with the subjective data, we show in Table 2 (first row) the corresponding PPCs. The p-values are all $p < 0.001$. From the comparison, we observe that LC_{RS1} outperforms both FC and SE .

To better analyze the results, we report in Table 3 (first row) the correlation performance, expressed in terms of PCCs, of each single measure. The p-values are all $p < 0.001$ except for \mathcal{M}_9 and \mathcal{M}_{11} where $p < 0.1$. The results show that our proposal outperforms each single metric and confirm our initial hypothesis that a pool of measures can better predict image complexity perception.

The verbal descriptions recorded during Experiment 1 were mapped into a list of criteria that aggregates concepts with the same meaning. We summarize in Table 4 the most common criteria used, in terms of their frequency with respect to the observers. We underline that each observer could have used more than one criteria. The *quantity of objects*, *details* and *colors* are the criteria that seem to dominate the complexity perception in Experiment 1. Moreover the most frequent criteria *quantity of objects* of the verbal descriptions is in accordance with the highest coefficient a_8 obtained with the PSO.

Experiment 2

Experiment 2 is used to test the linear combination LC_{RS1} . The 49 images here considered (RS2 dataset), depict real world scenes. As in Experiment 1, the subjective data is processed to obtain the mean scores. The RS2 images, ordered with respect to increasing mean scores, are reported in Fig 5. Image on the top left corresponds to the minimum mean score, while image on the bottom right is the one with the highest one.

The scatter plot between mean scores and our complexity measure (LC_{RS1}) is reported in Fig 6. The monotone function that best fits the data is also shown with a continuous line. The scatter plots between mean scores and FC and SE respectively are also included in the Figure for a comparison. In Table 2 (second row) the corresponding PPCs are reported. The p-

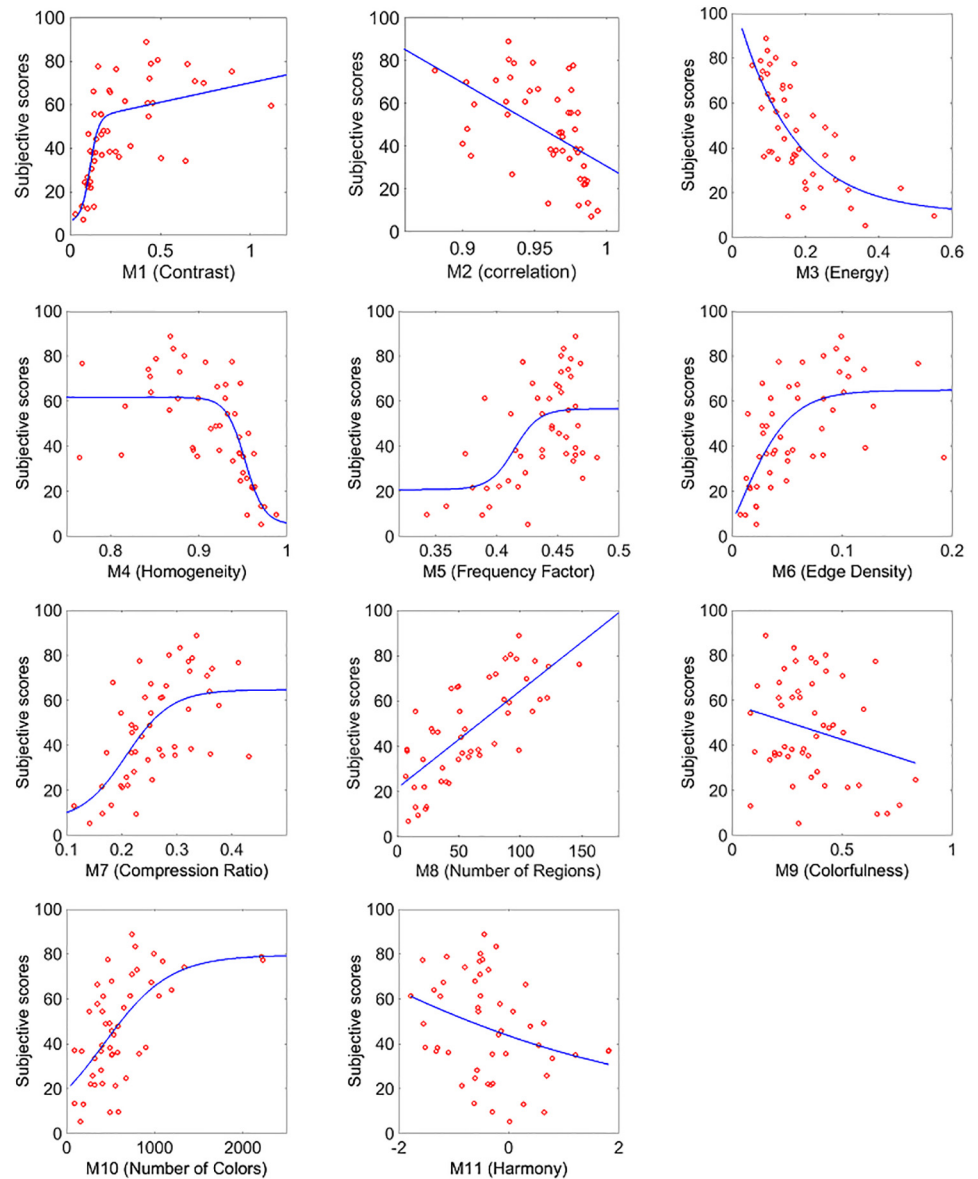


Fig 3. Scatter plots between mean scores and each of the single 11 measures for RS1 dataset.

doi:10.1371/journal.pone.0157986.g003

values are all $p < 0.001$. From the comparison, we observe that LC_{RS1} still outperforms both FC and SE .

In Fig 7 the scatter plots and the best fitting functions between each single measure and the mean scores of Experiment 2 are shown. The correlation performance in terms of PCCs is reported in Table 3 (second row). We observe that in general the performances are decreased with respect to the previous case (RS1 dataset). Measure \mathcal{M}_8 (Number of Regions) confirms to be among those with higher performance, together with \mathcal{M}_7 (Compression Ratio). Measures \mathcal{M}_9 and \mathcal{M}_{11} , developed mainly to evaluate color properties, show the lowest correlations with p-values $p < 0.1$. For the remaining metrics the p-values are all $p < 0.001$.

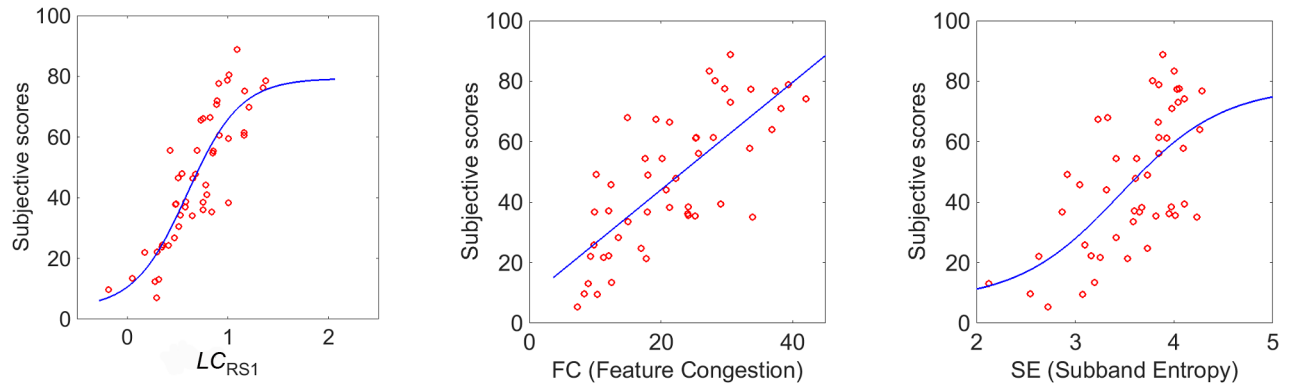


Fig 4. Scatter plots between mean scores and LC_{RS1} , FC and SE measures.

doi:10.1371/journal.pone.0157986.g004

Table 2. Correlation performance in terms of PCCs for LC_{RS1} , Feature Congestion and Subband Entropy measures for RS1, RS2, TXT1 and TXT2 datasets.

Measure	LC_{RS1}	Feature Congestion (FC)	Subband Entropy (SE)
PCC-RS1	0.86	0.75	0.64
PCC-RS2	0.81	0.74	0.66
PCC-TXT1	0.36	0.55	0.44
PCC-TXT2	0.66	0.71	0.15

doi:10.1371/journal.pone.0157986.t002

Table 3. Correlation performance in terms of PCCs for each of the single measures for RS1, RS2, TXT1 and TXT2 datasets.

Measure	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4	\mathcal{M}_5	\mathcal{M}_6	\mathcal{M}_7	\mathcal{M}_8	\mathcal{M}_9	\mathcal{M}_{10}	\mathcal{M}_{11}
PCC-RS1	0.75	0.52	0.72	0.56	0.57	0.67	0.65	0.74	0.25	0.63	0.30
PCC-RS2	0.64	0.47	0.56	0.64	0.45	0.66	0.67	0.69	0.35	0.43	0.42
PCC-TXT1	0.43	0.19	0.53	0.42	0.35	0.58	0.50	0.47	0.24	0.44	0.14
PCC-TXT2	0.64	0.66	0.77	0.64	0.57	0.47	0.59	0.62	0.14	0.37	0.37

doi:10.1371/journal.pone.0157986.t003

Table 4. Summary of verbal descriptions corresponding to Experiment 1.

Criterion	Frequency
Quantity of objects	48%
Quantity of details	33%
Quantity of colors	24%
Order and regularity	19%
Understandability	9%

doi:10.1371/journal.pone.0157986.t004

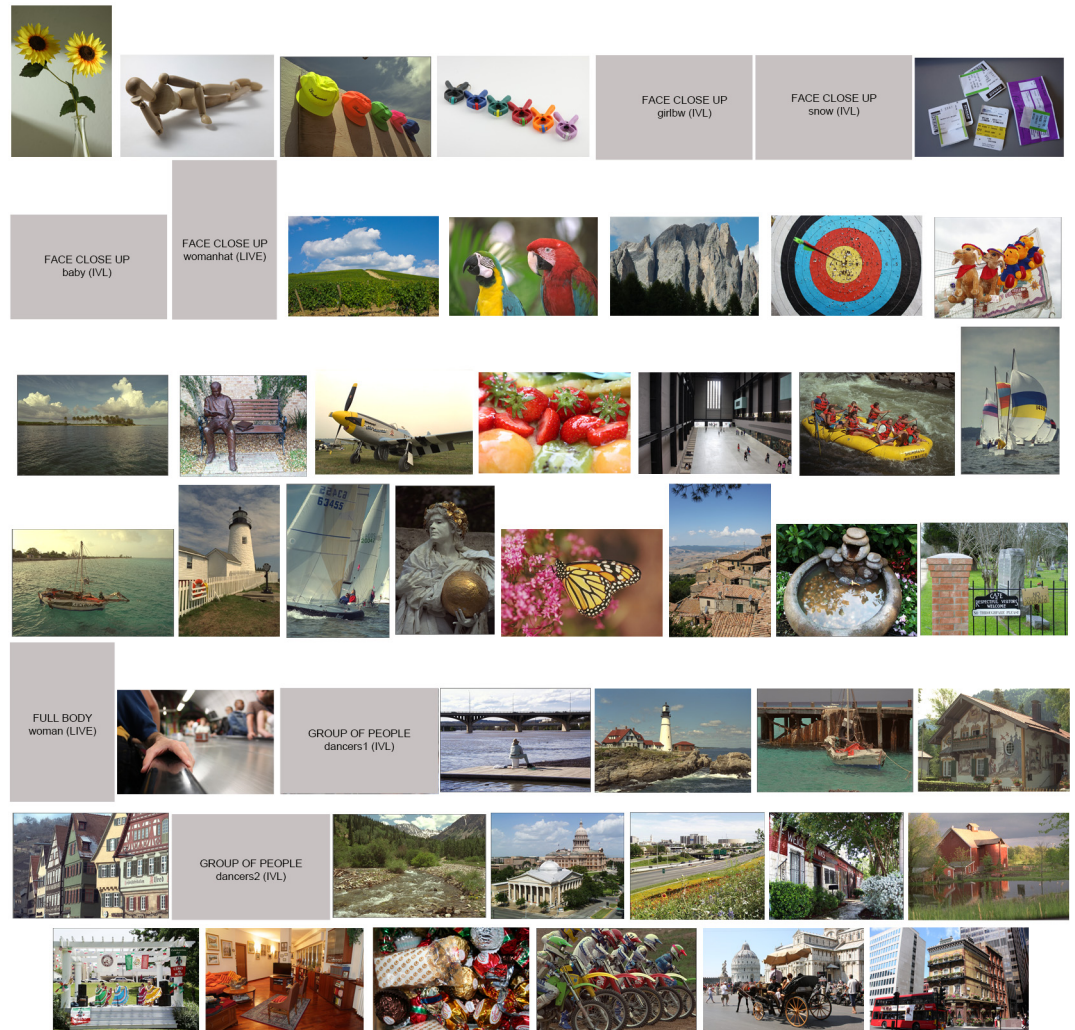


Fig 5. Ranking of the 49 images of the RS2 dataset according to increasing mean scores. Lowest mean score = 19 top left, highest mean score = 74, bottom right. Images with faces or recognizable people are not depicted, but are labeled with the name in the corresponding database.

doi:10.1371/journal.pone.0157986.g005

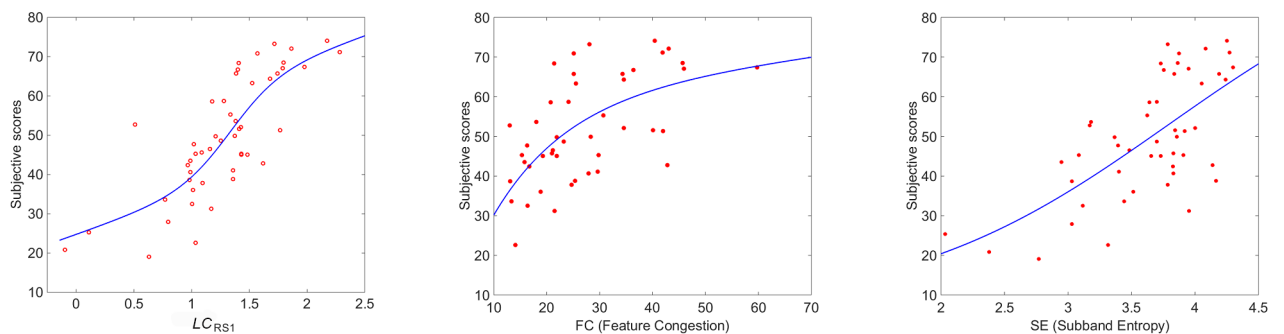


Fig 6. Scatter plots between subjective sscores and LC_{RS1} , FC and SE measures, respectively on RS2.

doi:10.1371/journal.pone.0157986.g006

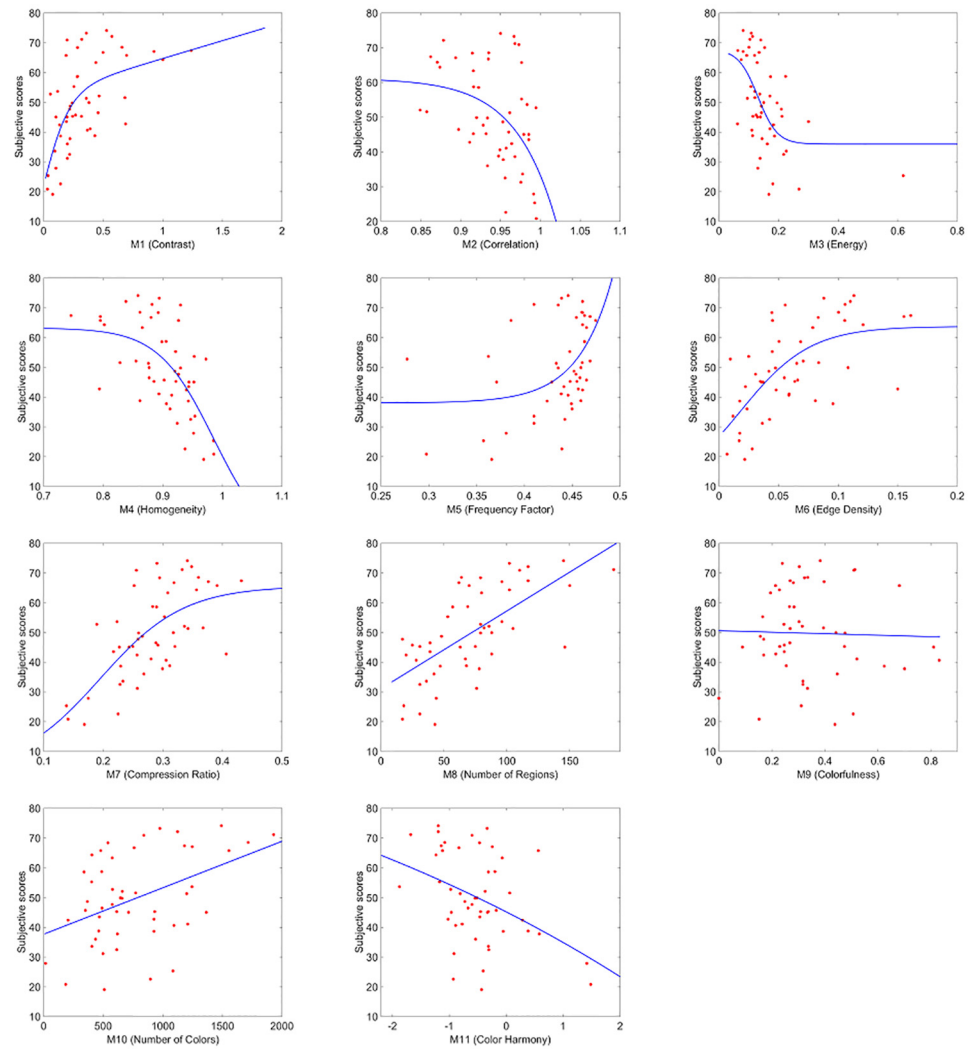


Fig 7. Scatter plots between mean scores and the single measures in the case of RS2 dataset.

doi:10.1371/journal.pone.0157986.g007

Table 5 summarizes the most common criteria collected during Experiment 2. From the analysis of the table we confirm the results of Experiment 1. As in the first Experiment the most relevant criteria used during the complexity assessment are *quantity of objects, details and colors*. They are also adopted with similar frequencies by the observers. Two more criteria are also used: *familiarity and texture*.

Experiment 3

In Experiment 3 we investigate how the complexity measure LC_{RS1} behaves when a different kind of stimuli is used. We thus consider as stimuli the 54 texture images of TXT1 dataset. Mean scores are obtained as in Experiments 1 and 2. In Fig 8 stimuli are shown in increasing order of complexity, according to mean scores. Image on the top left corresponds to the minimum mean score, while image on the bottom right has the highest one.

Table 5. Summary of verbal descriptions corresponding to Experiment 2.

Criterion	Frequency
Quantity of objects	53%
Quantity of details	31%
Quantity of colors	31%
Understandability	22%
Order and regularity	19%
Familiarity	16%
Texture	12%

doi:10.1371/journal.pone.0157986.t005



Fig 8. Texture images of TXT1 dataset, ordered from the less complex (top left) to the most complex (bottom right), according to the mean scores.

doi:10.1371/journal.pone.0157986.g008

We therefore correlate the linear combination LC_{RS1} with the subjective data collected for the texture stimuli and we find that it does not perform well: $PCC = 0.36$ with $p = 0.001$. For a comparison, in Table 2 (third row) we report also the PPCs corresponding to FC and SE applied to the TXT dataset. The p-values are all $p < 0.001$. We observe that also for FC and SE the correlation performance is decreased with respect to the case of real world scenes datasets.

Given the different kind of stimuli here used (texture vs real scenes) we propose to tune the weighting coefficients of the linear combination on this new set of stimuli. As before, we have performed 1000 runs of the PSO to obtain the new set of parameters. Within the 1000 runs, the average PCC results equal to 0.79 with standard deviation equal to 0.02 (minimum PCC = 0.67, maximum PCC = 0.83). In Table 6 we report the 11 weighting coefficients averaged over the 1000 run. We call the linear combination obtained using these coefficients LC_{TXT1} .

Comparing Tables 1 and 6, we observe that the linear combinations LC_{RS1} and LC_{TXT1} reflect the different nature of the stimuli (RS versus TXT) in the sign and absolute values of the coefficients. The highest contribution to the linear combination in LC_{TXT1} comes from \mathcal{M}_5

Table 6. Weighting coefficients for the linear combination LC_{TXT1} , obtained after averaging the parameters over 1000 runs of PSO.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
0.3653	0.0913	-0.0161	0.0127	-0.7150	0.3961	0.2774	-0.0358	-0.0099	0.4854	0.0209

doi:10.1371/journal.pone.0157986.t006

(Frequency Factor) followed by M_{10} (Number of Colors) and M_6 (Edge Density). The correlation coefficient PCC between LC_{TXT1} and the mean scores is now increased and is equal to 0.81. The scatter plots between the mean scores of Experiment 3 and LC_{TXT1} , FC and SE as well as the best fit of the data are shown in Fig 9 respectively. In Table 7 we summarize the results reporting the PCCs of LC_{TXT1} , LC_{RS1} , FC , and SE applied to TXT1 dataset. The p-values are all $p < 0.001$. We observe that the linear combination tuned on the texture set of stimuli outperforms all the others.

We also correlate the subjective data collected for the TXT1 dataset with the 11 single measures, reporting in Table 2 (third row) their performance. We observe that in general the single measures do not perform very well. Moreover, for measures M_2 (Correlation) and M_{11} (Color Harmony) we were not able to find a significant correlation. Only three of them show PCC greater or equal to 0.5 with p-values $p < 0.001$. These three measures are: M_3 (Energy), M_6 (Edge Density) and M_7 (Compression Ratio). For the remaining measures the p-values are $p < 0.01$.

We summarize in Table 8 the verbal descriptions of the observers. With this kind of stimuli the most frequent criteria adopted are regularity, understandability, quantity of details and familiarity, in agreement with the results of Guo et al. [34].

Experiment 4

Experiment 4 is used to further test the linear combination LC_{TXT1} . The 58 images of TXT2 dataset used in Experiment 4 are reported in increasing order of complexity in Fig 10. Subjective scores of this experiment are used to test the linear combination LC_{TXT1} . Its correlation performance is shown in Table 9 and it is compared with the performance on the training set TXT1. Our results confirm that also for the case of texture stimuli, the linear combination proposed LC_{TXT1} outperforms on the test set all the single measures considered (11 measures, FC and SE) and the LC_{RS1} , as it can be seen from Tables 2 and 3.

Discussions

From our investigation two aspects of image complexity can be underlined. Many different perceptual properties are involved in image complexity evaluation and their relative influence

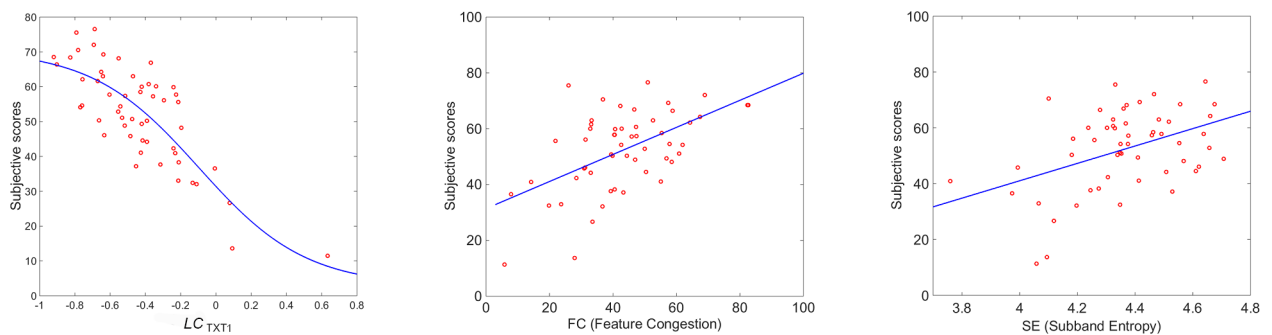


Fig 9. Scatter plots between mean scores of Experiment 3 and LC_{TXT1} , FC and SE respectively.

doi:10.1371/journal.pone.0157986.g009

Table 7. Correlation performance in terms of PCCs for LC_{TXT1} , FC and SE on TXT1 dataset.

Measure	LC_{TXT1}	LC_{RS1}	FC	SE
PCC	0.81	0.36	0.55	0.44

doi:10.1371/journal.pone.0157986.t007

Table 8. Summary of verbal descriptions corresponding to Experiment 3.

Criterion	Frequency
Regularity	60%
Understandability	47%
Quantity of details	33%
Familiarity	13%
Quantity of colors	12%

doi:10.1371/journal.pone.0157986.t008



Fig 10. Texture images of TXT2 dataset, ordered from the less complex (top left) to the most complex (bottom right), according to the mean scores.

doi:10.1371/journal.pone.0157986.g010

Table 9. Correlation performance in terms of PCCs for LC_{TXT1} , on TXT1 and TXT2 datasets.

LC_{TXT1}	TXT1	TXT2
PCC	0.81	0.81

doi:10.1371/journal.pone.0157986.t009

depends on the type of stimuli. These considerations are supported by both our computational proposal and the analysis of the verbal descriptions.

Analyzing the subjective results of all the four experiments we can try to extract some general considerations about image complexity perception. We separate the following analysis with respect to the different kind of stimuli used. In the case of real world scenes, from Figs 2 and 5 we observe that images with few objects and close-ups are judged as the less complex ones, while on the other hand buildings and streetscapes mainly belong to the most complex ones. These results are in agreement with those obtained by Purchase et al. [31], who addressed image complexity within the field of web interface design.

Analyzing the verbal descriptions reported by the observers while evaluating image complexity (Tables 4 and 5) we note that *quantity of objects*, *details* and *colors* are the criteria that seem to dominate the complexity perception in both Experiments 1 and 2. For Experiment 2, other two criteria are also used: *familiarity* and *texture*. We observe that many of these verbal descriptions agree with the different definitions found in the literature about image complexity and above reported (section Introduction): Snodgrass et al. [17] refer to the visual complexity as the *amount of details* in an image, Heaps and Hande [18] define complexity as the degree of difficulty in providing a verbal description of an image (*understandability*), Forsythe [20] refers that image complexity should be considered in relation to *familiarity*. We also note that similar criteria have been found in the study by Oliva et al. [30] who use indoor scenes as stimuli. In fact, the authors reported that the criteria corresponding to variety and quantity of objects and colors dominated the representation of complexity, followed by concepts like clutter, symmetry, open space, organization and contrast.

Trying to associate these verbal descriptions with the single objective measures, we can associate the criteria *quantity of objects*, *details* and *colors* to \mathcal{M}_8 (Number of Regions), \mathcal{M}_{10} (Number of Colors), and \mathcal{M}_6 (Edge Density) respectively. The description *order and regularity* can be in correspondence with the visual clutter measures *FC* and *SE*. While *quantity of objects*, *details* and *colors* and *order and regularity* can be associated to bottom-up cognitive mechanisms, *understandability* and *familiarity* that play also an important role, are clearly related to top-down processes and none of the considered measures alone is able to capture these concepts. Moreover, several observers have reported both types of criteria (bottom-up and top-down), confirming that bottom-up and top-down mechanisms interfere in perception.

Regarding the texture stimuli, from Figs 8 and 9 we can notice that images with regular pattern and symmetries have been judged as less complex, while images with more details and less ordered structures have been judged as more complex. These findings are in accordance with those obtained by Heaps and Handel [36]. They ranked the complexity of 24 texture images, printed in grayscale and belonging to the same VisTex database as ours and they observed that “textures with repetitive and uniform oriented patterns were judged less complex than disorganized patterns”.

The order of importance of the verbal descriptions of Experiment 3 has changed with respect to the corresponding ones of Experiments 1 and 2 (see Tables 4, 5 and 8). For texture images *regularity* is the most relevant criteria (reported by 60% of the observers), followed by *understandability* (47%). Instead for real world scenes these two criteria are among the ones less used (19% and 9% respectively in Experiment 1, and 19% and 22% in Experiment 2). These results are in accordance with those obtained by Yin et al. [25], who used sample images from Brodatz’s album [59]. They found that regularity, understandability, roughness, directionality, and density are the main characteristics that affect the visual complexity perception of texture images.

The differences between complexity perception of real scenes and texture patches are mainly related to the different image content and are reflected by the different order of importance

and frequency of the criteria reported in the verbal descriptions. In particular real scene images are more easily understandable than texture images. Analyzing the verbal descriptions reported by the observers we note that understandability is present for all the experiments. However in case of texture images it was more frequent, denoting that probably observers pay more attention on this aspect while evaluating texture complexity. Instead, in case of real world scenes understandability has been used with less frequency, as probably real scenes are intrinsically more understandable.

Comparing the performance of the single features, in terms of PCCs (see [Table 3](#)), we observe that in general those obtained for real world datasets (RS1 and RS2) are higher than the corresponding ones obtained for texture images. Following a similar trend, our linear combination proposal trained on real world scenes (LC_{RS1}) shows a significantly low performance applied to texture images (see [Table 2](#)). However, we have demonstrated that if the parameters of the linear combination are optimized with respect to this new dataset (TXT1), a significant improvement is reached. The final performance is thus comparable with that previously obtained on real world images.

Also Chikhman et al. [29] concluded that for different types of images, different measures of complexity may be required. In fact they found that for outline objects the best predictor of their experimental data was number of turns of an image, while for the hieroglyphs set of stimuli, the best correlation was given by the product of the square spatial frequency median and the image area. Focusing on streetscape images, Cavalcante et al. [32] proposed to combine contrast and spatial frequency to form a single objective measure. Within the same kind of streetscape images they found that their proposal is more effective and robust for nighttime scenes than for daytime ones and thus they concluded that “objective measures based on reduced sets of low-level image characteristics are unlikely to be satisfactory for all possible streetscapes”.

To assess complexity of painting images, Guo et al. [37] proposed to combine several visual features in a machine learning approach to classify images into three groups of high, medium and low complexity. In this way the authors were able to achieve a high classification performance. Moreover, studying the role of the single feature, they found that features related to hue and local color variations mostly conditioned the classification results in the specific case of painting images.

In our study we have considered real world images that cover several contents, including textures. We have considered eleven visual features, but our computational proposal can be extended either using different features or increasing the number of them, to cope with other different kind of stimuli.

Finally we remark that in our work we have considered databases of high quality images. An interesting aspect that should be investigated in the future is the interference between image quality and complexity perception: how do different kind of distortions influence image complexity perception? How does image complexity influence quality perception? The results of this analysis in fact could give an important insight in interpreting how signals and artifacts mutually interfere while evaluating image complexity.

Conclusions

Real world images show a high variability in depicted content. When observer are asked to assess image complexity, it emerges that complexity perception is guided by different criteria. Some of them are related to visual features representing bottom-up processes, while others are more related to top-down ones. Moreover, several criteria are adopted simultaneously by each observer, showing a multidimensional aspect of complexity. Thus in this work we have

proposed a complexity measure that linearly combines several visual features. When this measure is applied to real world scenes and texture it is able to predict the complexity perception outperforming each single feature. The weighting coefficients of the linear combination depend on the kind of stimuli and the relative contribution of the measures mostly reflects the criteria used by the observers. As a future work, to define a more general complexity measure, we are planning to mix images belonging to real scene and texture databases on which conduct new psycho-physical experiments. Furthermore, we plan to investigate non-linear models to combine the single measures, as for example Support Vector Machine or Genetic Programming to take into account different modalities of interactions of visual properties.

Supporting Information

S1 File. Objective measures and subjective scores of images in the RS1 real world scenes database.

(XLS)

S2 File. Objective measures and subjective scores of images in the RS2 real world scenes database.

(XLS)

S3 File. Objective measures and subjective scores of images in the TXT1 texture images database.

(XLS)

S4 File. Objective measures and subjective scores of images in the TXT2 texture images database.

(XLS)

Acknowledgments

We acknowledge the Laboratory for Image and Video Engineering and Center for Perceptual Systems at the University of Texas at Austin, the Imaging and Vision Laboratory and the Department of Informatics Systems and Communication at the University of Milano Bicocca, and the Vision and Modeling Group of the MIT Media Laboratory for providing the image databases.

Author Contributions

Conceived and designed the experiments: SC GC EB FG. Performed the experiments: SC GC EB FG. Analyzed the data: SC GC EB FG. Contributed reagents/materials/analysis tools: SC GC EB FG. Wrote the paper: SC GC EB FG.

References

1. Forsythe A, Sheehy N, Sawey M. Measuring icon complexity: An automated analysis. *Behavior Research Methods, Instruments, & Computers*. 2003; 35(2):334–342. doi: [10.3758/BF03202562](https://doi.org/10.3758/BF03202562)
2. Reinecke K, Yeh T, Miratrix L, Mardiko R, Zhao Y, Liu J, et al. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM; 2013. p. 2049–2058.
3. Ramanarayanan G, Bala K, Ferwerda JA, Walter B. Dimensionality of visual complexity in computer graphics scenes. In: *Electronic Imaging 2008*. International Society for Optics and Photonics; 2008. p. 68060E–68060E.
4. Wei B, Guan T, Duan L, Yu J, Mao T. Wide area localization and tracking on camera phones for mobile augmented reality systems. *Multimedia Systems*. 2015; 21(4):381–399. doi: [10.1007/s00530-014-0364-2](https://doi.org/10.1007/s00530-014-0364-2)

5. Yaghmaee F, Jamzad M. Estimating watermarking capacity in gray scale images based on image complexity. *EURASIP Journal on Advances in Signal Processing*. 2010; 2010:8. doi: [10.1155/2010/851920](https://doi.org/10.1155/2010/851920)
6. Perkiö J, Hyvärinen A. Modelling image complexity by independent component analysis, with application to content-based image retrieval. In: *Artificial Neural Networks—ICANN 2009*. Springer; 2009. p. 704–714.
7. Guan T, Wang Y, Duan L, Ji R. On-Device Mobile Landmark Recognition Using Binarized Descriptor with Multifeature Fusion. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2015; 7(1):12.
8. Zhang Y, Guan T, Duan L, Wei B, Gao J, Mao T. Inertial sensors supported visual descriptors encoding and geometric verification for mobile visual location recognition applications. *Signal Processing*. 2015; 112:17–26. doi: [10.1016/j.sigpro.2014.08.029](https://doi.org/10.1016/j.sigpro.2014.08.029)
9. Wei B, Guan T, Yu J. Projected residual vector quantization for ANN search. *MultiMedia, IEEE*. 2014; 21(3):41–51. doi: [10.1109/MMUL.2013.65](https://doi.org/10.1109/MMUL.2013.65)
10. Huahui Wu MC, Kinicki R. A study of video motion and scene complexity. In: *Tech. Rep. WPI-CS-TR-06–19*, Worcester Polytechnic Institute; 2006.
11. Peters RA, Strickland RN. Image complexity metrics for automatic target recognizers. In: *Automatic Target Recognizer System and Technology Conference*; 1990. p. 1–17.
12. Xu J, Jiang M, Wang S, Kankanhalli MS, Zhao Q. Predicting human gaze beyond pixels. *Journal of vision*. 2014; 14(1):28. doi: [10.1167/14.1.28](https://doi.org/10.1167/14.1.28) PMID: [24474825](https://pubmed.ncbi.nlm.nih.gov/24474825/)
13. Zhao Q, Koch C. Advances in Learning Visual Saliency: From Image Primitives to Semantic Contents. In: *Neural Computation, Neural Devices, and Neural Prosthesis*. Springer; 2014. p. 335–360.
14. Donderi DC. Visual complexity: a review. *Psychological Bulletin*. 2006; 132(1):73. doi: [10.1037/0033-2909.132.1.73](https://doi.org/10.1037/0033-2909.132.1.73) PMID: [16435958](https://pubmed.ncbi.nlm.nih.gov/16435958/)
15. Marin MM, Leder H. Examining Complexity across Domains: Relating Subjective and Objective Measures of Affective Environmental Scenes, Paintings and Music. *PLoS ONE*. 2013; 8(8):e72412. doi: [10.1371/journal.pone.0072412](https://doi.org/10.1371/journal.pone.0072412) PMID: [23977295](https://pubmed.ncbi.nlm.nih.gov/23977295/)
16. Kolmogorov AN. Three approaches to the quantitative definition of information'. *Problems of information transmission*. 1965; 1(1):1–7.
17. Snodgrass JG, 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*. 1980; 6(2):174.
18. Heaps C, Handel S. Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*. 1999; 25(2):299.
19. Birkhoff GD. *Collected mathematical papers*. New York; 1950.
20. Forsythe A. Visual Complexity: Is That All There Is? In: *Engineering Psychology and Cognitive Ergonomics*. Springer; 2009. p. 158–166.
21. Cardaci M, Di Gesù V, Petrou M, Tabacchi ME. On the evaluation of images complexity: A fuzzy approach. In: *Fuzzy Logic and Applications*. Springer; 2006. p. 305–311.
22. Palumbo L, Ogden R, Makin AD, Bertamini M. Examining visual complexity and its influence on perceived duration. *Journal of vision*. 2014; 14(14):3. doi: [10.1167/14.14.3](https://doi.org/10.1167/14.14.3) PMID: [25487112](https://pubmed.ncbi.nlm.nih.gov/25487112/)
23. Chacon MIM, Aguilar LED, Delgado AS. Fuzzy adaptive edge definition based on the complexity of the image. In: *10 th IEEE International Conference on Fuzzy Systems*; 2001. p. 675–678.
24. Mario I, Chacon M, Alma D, Corral S. Image complexity measure: a human criterion free approach. In: *Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American. IEEE*; 2005. p. 241–246.
25. Yin K, Wang L, Guo Y. Fusing Multiple Visual Features for Image Complexity Evaluation. In: *Advances in Multimedia Information Processing—PCM 2013*. Springer; 2013. p. 308–317.
26. 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*. Eurographics Association; 2005. p. 177–184.
27. Rosenholtz R, Li Y, Nakano L. Measuring visual clutter. *Journal of Vision*. 2007; 7(2):17. doi: [10.1167/7.2.17](https://doi.org/10.1167/7.2.17) PMID: [18217832](https://pubmed.ncbi.nlm.nih.gov/18217832/)
28. Mack M, Oliva A. Computational estimation of visual complexity. In: *the 12th Annual Object, Perception, Attention, and Memory Conference*; 2004.
29. Chikhman V, Bondarko V, Danilova M, Goluzina A, Shelepin Y. Complexity of images: Experimental and computational estimates compared. *Perception*. 2012; 41:631–647. doi: [10.1068/p6987](https://doi.org/10.1068/p6987) PMID: [23094454](https://pubmed.ncbi.nlm.nih.gov/23094454/)

30. Oliva A, Mack ML, Shrestha M. Identifying the Perceptual Dimensions of Visual Complexity of Scenes. In: Proc. 26th Annual Meeting of the Cognitive Science Society; 2004.
31. Purchase HC, Freeman E, Hamer J. Predicting Visual Complexity. In: Proceedings of the 3rd International Conference on Appearance, Edinburgh, UK; 2012. p. 62–65.
32. Cavalcante A, Mansouri A, Kacha L, Barros AK, Takeuchi Y, Matsumoto N, et al. Measuring street-scape complexity based on the statistics of local contrast and spatial frequency. PLoS ONE. 2014; 9 (2). doi: [10.1371/journal.pone.0087097](https://doi.org/10.1371/journal.pone.0087097)
33. Ciocca G, Corchs S, Gasparini F, Bricolo E, Tebano R. Does color influence image complexity perception? In: Fifth IAPR Computational Color Imaging Workshop (CCIW'15). vol. 9016 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg; 2015. p. 139–148.
34. Guo X, Asano CM, Asano A, Kurita T, Li L. Analysis of texture characteristics associated with visual complexity perception. Optical review. 2012; 19(5):306–314. doi: [10.1007/s10043-012-0047-1](https://doi.org/10.1007/s10043-012-0047-1)
35. Ciocca G, Corchs S, Gasparini F. Complexity Perception of Texture Images. In: Murino V, Puppo E, Sona D, Cristani M, Sansone C, editors. New Trends in Image Analysis and Processing—ICIAP 2015 Workshops. vol. 9281 of Lecture Notes in Computer Science. Springer International Publishing; 2015. p. 119–126.
36. Heaps C, Handel S. Similarity and features of natural textures. Journal of Experimental Psychology: Human Perception and Performance. 1999; 25(2):299.
37. Guo X, Kurita T, Asano CM, Asano A. Visual complexity assessment of painting images. In: Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE; 2013. p. 388–392.
38. J K, R E. Particle swarm optimization. In: Proc IEEE Int Conf Neural Networks. vol. 4; 1995. p. 1942–1948.
39. Bianco S, Schettini R. Two new von Kries based chromatic adaptation transforms found by numerical optimization. Color Research & Application. 2010; 35(3):184–192. doi: [10.1002/col.20573](https://doi.org/10.1002/col.20573)
40. Yu CP, Samaras D, Zelinsky GJ. Modeling visual clutter perception using proto-object segmentation. Journal of vision. 2014; 14(7):4. doi: [10.1167/14.7.4](https://doi.org/10.1167/14.7.4) PMID: [24904121](https://pubmed.ncbi.nlm.nih.gov/24904121/)
41. Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca, <http://www.ivl.disco.unimib.it/activities/complexity-perception-in-images>; 2016
42. Sheik H, Wang Z, Cormack L, Bovik A. In: LIVE Image Quality Assessment Database Release 2. <http://live.ece.utexas.edu/research/quality>; 2006.
43. Sheikh HR, Sabir MF, Bovik AC. A statistical evaluation of recent full reference image quality assessment algorithms. Image Processing, IEEE Transactions on. 2006; 15(11):3440–3451. doi: [10.1109/TIP.2006.881959](https://doi.org/10.1109/TIP.2006.881959)
44. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on. 2004; 13(4):600–612. doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861)
45. Corchs S, Gasparini F, Schettini R. No Reference Image Quality classification for JPEG-Distorted Images. Digital Signal Processing. 2014; 30:86–100. doi: [10.1016/j.dsp.2014.04.003](https://doi.org/10.1016/j.dsp.2014.04.003)
46. Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca, <http://www.ivl.disco.unimib.it/activities/image-quality>; 2014
47. MIT Media Lab, Vision texture homepage, <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>.
48. Cusano C, Napoletano P, Schettini R. Evaluating color texture descriptors under large variations of controlled lighting conditions. Journal of the Optical Society of America A. 2016 1; 33(1):17–30. doi: [10.1364/JOSAA.33.000017](https://doi.org/10.1364/JOSAA.33.000017)
49. Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca, <http://projects.ivl.disco.unimib.it/rawfoot>; 2016.
50. Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca, <http://projects.ivl.disco.unimib.it/>; 2016.
51. Freedman DA. Statistical models: theory and practice. Cambridge University Press; 2009.
52. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on. 1973;(6):610–621. doi: [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314)
53. Corchs S, Gasparini F, Schettini R. Grouping strategies to improve the correlation between subjective and objective image quality data. In: IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics; 2013. p. 86530D–86530D.
54. Corchs S, Gasparini F, Schettini R. No reference image quality classification for JPEG-distorted images. Digital Signal Processing. 2014; 30:86–100. doi: [10.1016/j.dsp.2014.04.003](https://doi.org/10.1016/j.dsp.2014.04.003)

55. Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2002; 24(5):603–619. doi: [10.1109/34.1000236](https://doi.org/10.1109/34.1000236)
56. Hasler D, Suesstrunk SE. Measuring colorfulness in natural images. In: *Electronic Imaging 2003*. International Society for Optics and Photonics; 2003. p. 87–95.
57. Artese MT, Ciocca G, Gagliardi I. Good 50x70 Project: A portal for Cultural And Social Campaigns. In: *Archiving Conference*. vol. 2014. Society for Imaging Science and Technology; 2014. p. 213–218.
58. Solli M, Lenz R. Color harmony for image indexing. In: *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE; 2009. p. 1885–1892.
59. Brodatz P. *Textures: a photographic album for artists and designers*. Dover Pubns; 1966.