A Genetic Programming approach to evaluate complexity of texture images

Gianluigi Ciocca[a,b], **Silvia Corchs**[a,b,*], **Francesca Gasparini**[a,b]
[a]Dipartimento di Informatica, Sistemistica e Comunicazione, University of Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy
[b]NeuroMi - Milan Center for Neuroscience - Italy

Abstract. In this paper we adopt Genetic Programming (GP) to define a measure that can predict complexity perception of texture images. We perform psycho-physical experiments on three different datasets to collect data on the perceived complexity. The subjective data are used for training, validation, and test of the proposed measure. These data are also used to evaluate several possible candidate measures of texture complexity related to both low level and high level image features. We select four of them (namely roughness, number of regions, chroma variance, and memorability) to be combined in a GP framework. This approach allows a non-linear combination of the measures and could give hints on how the related image features interact in complexity perception. The proposed complexity measure $\mathcal{M}_{\mathcal{GP}}$ exhibits Pearson Correlation Coefficients of 0.890 on the training set, 0.728 on the validation set, and 0.724 on the test set. $\mathcal{M}_{\mathcal{GP}}$ outperforms each of all the single measures considered. From the statistical analysis of different GP candidate solutions, we found that the roughness measure evaluated on the gray level image is the most dominant one, followed by the memorability, the number of regions, and finally the chroma variance.

Keywords: Texture, image complexity, Genetic Programming, image features.

*Silvia Corchs, silvia.corchs@disco.unimib.it

1 Introduction

Visual scenes are composed of numerous textures, objects, and colors. Texture helps us to understand the visual world. A strict definition of texture is difficult. It can be regarded as what constitutes a macroscopic region, with repetitive patterns in which elements are arranged according to a placement rule.[1] Texture provides a cue to the shape and orientation of a surface, to segmenting an image into meaningful regions, and to classifying those regions in terms of material properties.[2] Human texture processing has not yet been fully understood given its complexity and the involvement of mechanisms at different levels. Researches have addressed the problem of texture processing using both artificial and natural materials, mainly in the field of texture classification.[3] In this work we address the analysis of texture images from a different point of view, focusing on human perception. In particular we investigate the perceived complexity of natural texture images. Also Tamura et al.[4] have approached texture perception, not with respect to complexity, but with the different goal of defining computational measures to describe texture. They set up an experiment where subjects evaluated texture images in a pair-wise comparison, with respect to six textural features: coarseness, contrast, directionality, line-likeness, regularity and roughness. They thus defined computational measures of these features that correlate with the experimental results. How humans perceive texture can provide new insights in understanding the process of material recognition. In particular, the study of image complexity perception can be useful in many different domains. For example to provide an a priori estimate of the difficulty of locating a target in an image.[5] It can also be embedded in the similarity measure for content-based image retrieval.[6] In the context of human-computer interaction, complexity can be used to estimate the level of usability of icons[7] or websites.[8] It can also be exploited in watermarking algorithms to estimate the amount of information that can be hidden in images,[9] or used in image compression algorithms[10] to estimate the bandwidth allocation required. Prediction of saliency within images can take advantage of the definition of complexity measures.[11,12] Finally, the image complexity concept is also used by neuroscientists interested in the mechanisms of recognition, learning and memory.[13]

In previous works, image complexity is either evaluated using a single image feature or by an empirical selection and combination of different image features. But all the works agree in that different perceived aspect of an image intervene in the assessment of image complexity. However how to integrate them within a model that also take into account top-down mechanisms is still an open issue. In this work we try to derive a complexity measure by automatically combining different image feature measures within an optimization framework. We are interested in the computational model of the overall complexity measure, and in the insights that its formulation can give us for the understanding of texture image complexity. Here we address these issues using a Genetic Programming (GP) framework to design a measure to predict texture image complexity. GP permits to combine different measures of image features in a nonlinear way that can be more suitable to match subjective perception. We have adopted GP because it does not require an a-priori definition of the form of the solution, that can be linear as well as non-linear. Moreover it does not require to specify the size, and the shape, of the solution in advance.[14,15] As a further advantage, GP provides intelligible and easily interpretable solutions.

Therefore, in this work we have:

- Performed three psycho-physical experiments to assess complexity perception of texture images. These experiments differ in the dataset of stimuli used: RawFooT,[16] VisTex,[17] and a grayscale version of VisTex and are described in Section 3.1. All the experimental data collected are available at our website.

- Analyzed eleven measures of different image features as possible candidates to be adopted in the GP approach. Five of them were specifically designed to evaluate texture properties by Tamura et al.,[4] Section 3.2.

- Designed a GP framework to combine in a nonlinear way selected measures of image features, to obtain a measure of image complexity (Section 4).

- Trained the GP framework on RawFooT dataset, validated it on VistTex, and finally tested it on the grayscale VisTex. The results are reported and discussed in Section 5.

In the next section we provide a literature survey on image complexity and we summarize the GP optimization framework as well.

2 Literature Survey

2.1 Related work

Depending on the specific task and the application domain, different definitions of image complexity are possible. From a purely mathematical point of view, Kolmogorov[18] defines the complexity of an object as the length of the shortest program that can construct the object from basic elements, or description language. Some studies of visual complexity perception deal with real world images. Oliva et al.[19] represented visual complexity by a multi-dimensional space, that includes quantity of objects, clutter, openness, symmetry, organization and variety of colors. Snodgrass et al.[20] refer to the visual complexity as the amount of detail or intricacy in an image. Birkhoff[21] relates the image complexity to visual aesthetics. However, little research has been carried out into the visual complexity of texture images. Rao and Lohse[22] showed that the three attributes: repetitiveness, orientation and complexity can be helpful in order to identify and classify natural textures. Within their work they defined complexity as the degree of difficulty in providing a verbal description of the texture. Later, studying similarity and features of natural textures, Heaps and Hande,[23] found that the complexity of a texture can be estimated along a one dimensional axis representing the degree of perceivable structure. Recently, Guo et al.[24,25] have also considered the perception of texture complexity. They identified low-level characteristics that influence texture complexity perception: regularity, roughness, directionality, and density. Moreover they found that subject perception is not only related to concepts that can be described with objective measures, but also this perception can be influenced by top-down mechanisms. In particular from the verbal description collected during their experiments, understandability turns out to be one of the most frequent criteria used.

The problem of complexity perception of real world texture images has been previously investigated by Ciocca et al.[26] Specifically, experiments were performed to evaluate how different image features, such as color, edge, regions, and composition, correlate with the subjective data collected in proper psycho-physical experiments. The results of these experiments evidenced that different aspects should be considered when formulating a model to predict texture complexity, and that complexity perception depends on several image features that interact. The most frequent criteria adopted by the observers (17 subjects from University of Milano Bicocca) were regularity, understandability, familiarity and edge density, in accordance with those obtained by Guo et al.,[24,25] in an experiment where 30 observers from the Hiroshima University were involved. These results suggested that there is probably a general consensus and not an evident cultural or linguistic bias while evaluating texture complexity. However this aspect is still an open issue, especially when visual complexity of images with different semantic contents is considered.

2.2 GP background

GP is widely and successfully used in the field of pattern recognition[27,28] and information and image retrieval.[29,30] Recently, several authors adopted genetic programming to solve problems of texture classification and segmentation.[31–33] In GP the computer programs, i.e. the individuals of a population, are represented by trees, where the leaves, also called terminals, represent the constants and the independent variables of the problem. The arithmetic functions that combine the terminals correspond to the nodes of the tree. An example of an individual represented as a tree is shown in Fig. 1. In this example the tree corresponds to the expression $5 \times X_1 + sin(X_2) + X_3$, where the terminals are the three variables X_1, X_2, and X_3 and the constant 5, while the used functions are: plus ($+$), times ($\times$), and sin.

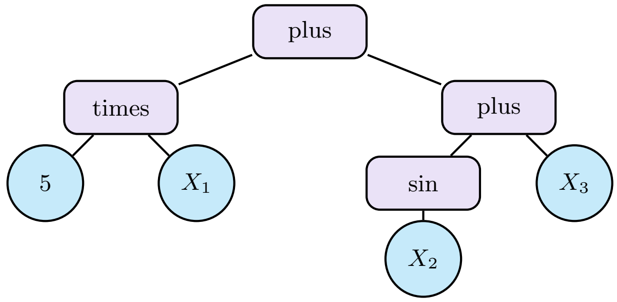

Fig 1 GP tree representing the program: $5 \times X_1 + sin(X_2) + X_3$

Initially a population of computer programs (individuals) are created randomly. A fitness function is designed with respect to the problem to be solved, and evaluated for each individual. The fitness function numerically quantifies the goodness of an individual computer program in solving that problem. The initial population, as in most evolutionary algorithms, evolves generation by generation. Individuals of a population are modified by applying genetic operators (such as reproduction, mutation, and crossover) to individuals selected with respect to their fitness. The reproduction operator selects the best individuals and copies them to the next generation. Mutation creates one new individual for the next population by randomly mutating a randomly chosen part of one selected individual. Crossover, instead, creates new individuals by recombining randomly chosen parts from two selected ones. These genetic operators are applied to obtain a new population, till an acceptable solution in terms of fitness is obtained or a stopping criteria is reached. The basic steps of Genetic Programming are summarized in Algorithm 1. For a deeper understanding of GP refer to consolidated works in the literature.[14,15]

Algorithm 1: Genetic Programming

begin
  Randomly create an initial Population P of individual computer programs;
  repeat
    Execute each program of the current Population P to evaluate its fitness;
    best-so-far individual $\leftarrow$ the best program in the population;
    Create a new empty population P';
    repeat
      Select one or two individual program(s) from Population P with respect to their fitness;
      Create new individuals for the Population P' by applying the genetic operators with specified probabilities on the selected individuals;
    until *P' is fully populated of individuals*;
    P $\leftarrow$ P';
  until *an acceptable solution in terms of fitness or a stopping criteria is found*;
  return *the best-so-far individual*
end

3 Complexity of texture images

3.1 Subjective data

To investigate the complexity perception of texture images we have setup a psycho-physical experiment, where texture images are individually shown on a web-interface. The images are shown in a random order, different for each subject. The subjects report their complexity judgments (scores) by dragging a slider onto a continuous scale in the range [0-100]. They can look at the stimuli for an unlimited time. The position of the slider is automatically reset after each evaluation. A grayscale chart is 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. Subjects that did not pass this test were not further considered. A limited subset of images have been used to make the subjects confident about the experimental procedure. The corresponding data collected have been discarded. All the participants who took part in the experiments were recruited in the Department of Informatics System and Communication of the University of Milano Bicocca. Even if some of them work in the field of image processing, no one is familiar with the research topic of image complexity. Each observer was taught about the experiment through the web-interface, where he/she also gave his/her informed consent.

With the described experimental setup, we have performed the following three experiments which differ for the dataset of texture images used as stimuli.

- **FOOD experiment**. In this experiment we have used texture images belonging to the Raw Food Texture dataset (RawFooT).[16] This dataset includes images of 68 samples of food textures, acquired under 46 lighting conditions, varying in light direction and intensity, illuminant color, or in a combination of these factors. The 68 classes of food depict various kind of meat, fish, cereals, fruit etc. The whole dataset includes 68 (classes) x 46 (lighting conditions) = 3128 images. In our work we have used as stimuli the subset of 68 food images taken under the D65 illuminant (i.e. daylight) and frontal light direction. Ten images were used in the preliminary phase of subject training, while the remaining 58 images were used in the actual experiment. A group of 25 observers took part of this experiment, among them 3 were discarded because they did not pass the Ishihara test. Of the 23 remained after the control test, 10 were women, mean age 28, range 19-68.

- **VisCol experiment**. In this experiment we have used texture images belonging to the VisTex dataset.[17] This dataset consists of 864 images representing 54 classes of natural objects or scenes captured under non-controlled conditions with a variety of devices. In our experiment we have used as stimuli 54 images, each one representative of the corresponding 54 classes, 9 further images were chosen for the training phase among the remaining ones. A group of 24 observers took part of this experiment, and only one of them was removed. Of the 23 remained after the control test, 7 were women, mean age 35, range 20-68.

- **VisGray experiment**. In this experiment we have used as stimuli the grayscale versions of the same texture images adopted in the VisCol experiment. All the 23 observers involved in this experiment passed the Ishihara test and thus they were all considered as valid subjects, 5 were women, mean age 30, range 18-65.

For each experiment, before averaging raw data across subjects we have processed them applying Z-score normalization[34] and outlier removal. This normalization permits to minimize the variation between the individual scores. This variation is mainly due to the fact that not all subjects use the full range of the numerical scale in evaluating images. The raw complexity score r_{ij} for the i-th subject and j-th image was first converted into Z-score z_{ij}:

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

where $\bar{r_i}$ is the average of the complexity scores over all images evaluated by the subject, and σ_i is the standard deviation. The Z-scores are then re-scaled in the range [0,100] and averaged across subjects $i = 1, ...S$ to obtain the average score y_j:[35]

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

The images of the three datasets used as stimuli are shown in Figs. 2-4 respectively. The images are reported in increasing order of complexity with respect to the average score. Images evaluated as the less complex are reported in the top left, while images with the highest complexity are reported in the bottom right.

Fig 2 Thumbnails of the texture images chosen to sample 58 classes of the RawFooT dataset, row-wise ordered from the less complex to the most complex, according to the Average Score.

3.2 Candidate complexity measures for texture images

We here considered several image features summarized in Table 1. Features from $\mathcal{M}_1$ to $\mathcal{M}_5$ were chosen among those frequently used in the literature to describe texture patterns.[4] $\mathcal{M}_6$, $\mathcal{M}_7$, and $\mathcal{M}_8$, also available in the literature, are related with image complexity perception as it was demonstrated by several works.[36–39] Since we are considering image complexity of both color and grayscale images, we have also introduced $\mathcal{M}_9$ and $\mathcal{M}_{10}$, that are image features mainly devoted to evaluate color image properties. All these candidate measures describe low level image features. We also aim to take into account high level concepts like understandability or familiarity. These two criteria were hinted from the analysis of the verbal descriptions adopted by the observers while evaluating texture complexity in a previous work.[26] This is in accordance with the experimental

Fig 3 Thumbnails of the texture images chosen to sample each of the 54 classes in the VisTex dataset, row-wise ordered from the less complex to the most complex), according to the Average Score.

results by Guo et al.[25] and Heaps and Handel[40] who considered complexity as the degree of difficulty in providing a verbal description of an image. We include the memorability index $\mathcal{M}_{11}$ proposed by Isola et al.[41] as a possible image feature of these two high level criteria.

To find out how the objective measures described in Table 1 can predict subjective scores, we have correlated each of them to the average score of the three experiments using a logistic regression function. The correlation performance is expressed in terms of the Pearson Correlation Coefficient (PCC) and reported in Table 2. For all the three datasets the best correlation performance is obtained with $\mathcal{M}_2$ and $\mathcal{M}_5$, both evaluated on the grayscale image. A high performance is also reached by $\mathcal{M}_9$ on the color datasets. With respect to all the remaining measures, only $\mathcal{M}_8$ shows correlation greater than 0.5 for all the three datasets considered.

Table 1 List of image features

	Feature	Brief description	Reference
$\mathcal{M}_1$	Coarseness	It relates to distances of notable spatial variations of gray levels. Implementation by Bianconi et al.[42]	Tamura et al.[4]
$\mathcal{M}_2$	Contrast	Ratio of the standard deviation and the kurtosis of the distribution of the gray levels. Implementation by Bianconi et al.[42]	Tamura et al.[4]
$\mathcal{M}_3$	Directionality	Obtained from the gray levels histogram of local edge probabilities against their directional angle. Implementation by Bianconi et al.[42]	Tamura et al.[4]
$\mathcal{M}_4$	Linelikeness	It is defined as an average coincidence of the edge directions in the gray levels. Implementation by Bianconi et al.[42]	Tamura et al.[4]
$\mathcal{M}_5$	Roughness	It is related to the standard deviation of the normalized gray levels. Implementation by Bianconi et al.[42]	Tamura et al.[4]
$\mathcal{M}_6$	Edge density	Calculated using the Canny edge detector applied to the grayscale image.	Mack et al.[36]
$\mathcal{M}_7$	Compression Ratio	Ratio of the image JPEG compressed with Q factor = 100 and the full size uncompressed image.	Corchs et al.[43]
$\mathcal{M}_8$	Number of regions	Calculated using the mean shift algorithm. It can be applied either to color or gray scale images.	Comaniciu and Meer.[44]
$\mathcal{M}_9$	Chroma variance	Harmonic mean of chroma values in the YCbCr color space.	
$\mathcal{M}_{10}$	Colorfullness	Linear combination of the mean and standard deviation of the pixel cloud in the color plane.	Hasler and Susstrunk[45]
$\mathcal{M}_{11}$	Memorability	Probability that a viewer will detect a repeat of the image within a stream of pictures. It can be applied either to color or gray scale images.	Isola et al.[41]

Table 2 PCC of the 11 objective measures of Table 1 for each of the three experiments.

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$	$\mathcal{M}_8$	$\mathcal{M}_9$	$\mathcal{M}_{10}$	$\mathcal{M}_{11}$
FOOD	0.410	0.854	0.415	0.427	0.862	0.187	0.596	0.617	0.828	0.047	0.426
VisCol	0.334	0.631	0.161	0.049	0.628	0.585	0.454	0.582	0.695	0.330	0.330
VisGray	0.185	0.655	0.190	0.127	0.646	0.391	0.525	0.587	0	0	0.322

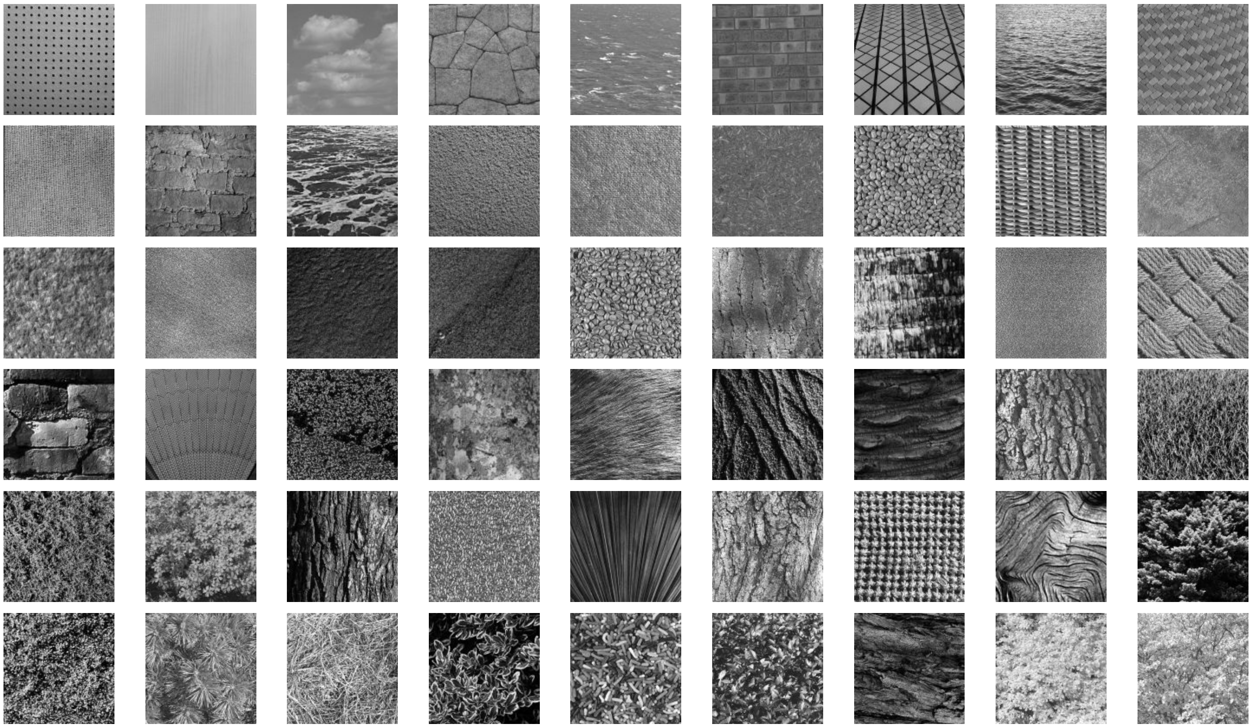

Fig 4 Thumbnails of the texture images chosen to sample each of the 54 classes in the VisTex dataset in their grayscale version, row-wise ordered from the less complex to the most complex, according to the Average Score.

4 Genetic Programming-based measure for image complexity

In this paper we design a measure to predict complexity of texture images using a Genetic Programming (GP) framework. GP is a domain-independent evolutionary method that genetically breeds a population of programs to solve a problem without requiring the user to know or specify the form or structure of the solution. In our context, GP permits to combine different measures in a nonlinear way that can be more suitable to match subjective perception of image complexity.

4.1 GP framework

In our work we have used the GPLAB toolbox, developed by Silva and Almeida.[46]

We here selected $\mathcal{M}_5$, $\mathcal{M}_8$, $\mathcal{M}_9$, $\mathcal{M}_{11}$ among the 11 measures listed in Table 1 as variables of the GP. $\mathcal{M}_5$ (Roughness), $\mathcal{M}_8$ (Number of regions), and $\mathcal{M}_9$ (Chroma variance) were chosen as they are the measures with the highest performance across the three datasets. We observe that

$\mathcal{M}_2$ (Contrast) is strictly correlated with $\mathcal{M}_5$ (see corresponding definitions in Table 1) and thus it was not considered. Furthermore, we have also chosen $\mathcal{M}_{11}$ even if its correlation is not so high to cope in some way with top-down effects that probably influence the complexity perception of texture images. We have here chosen as combining functions five different arithmetic operators, trying to describe the possible interactions of different visual properties: $+$ and $-$, to express the superposition principle, $\times$ and $\div$ to describe mutual reinforcement or attenuation, and log to describe eventual non linearity between objective evaluation and perception. 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 part of the fitness function. Starting with a set of training images $j = 1, ..N$, the candidate solution **X**=$\{x_j\}$ is previously transformed using a logistic function f,[47] to take into account the non linear mapping between objective and subjective data. The PCC of the candidate solution **X** is thus evaluated as follows:

$$PCC(\mathbf{X}) = \frac{\sum_{j=1}^{N}\left(f(x_j) - \overline{f(x)}\right)(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{N}\left(f(x_j) - \overline{f(x)}\right)^2}\sqrt{\sum_{j=1}^{N}(y_j - \bar{y})^2}} \tag{3}$$

where $f(x_j)$ is the logistically transformed value of the candidate solution for the j-th image, y_j is the corresponding average score (see Eq. 2) and $\overline{f(x)}$ and $\bar{y}$ are their averages over the selected training set.

As we have chosen only four variables, to penalize the GP solutions with a high number of nodes, we have introduced the following penalty term:

$$P(\mathbf{X}) = w * \frac{Number\_of\_terminal\_nodes(\mathbf{X})}{2 * Number\_of\_Variables(\mathbf{X})} \tag{4}$$

Table 3 Parameters used in the GP framework

Parameter	Description	Setting
Terminals	Leaves of the tree: variable and constants	$\mathcal{M}_5$, $\mathcal{M}_8$, $\mathcal{M}_9$, $\mathcal{M}_{11}$, 'rand(-1,1)'
Functions	Internal nodes: arithmetic operations	$+, -, \times, \div$, log
Fitness	Objective function to be optimized	Eq. (5)
Max number of generations	A stopping condition	25
Population size	Number of individuals (trees)	300
Max Tree depth	Size of the tree	Dynamic
Initialization algorithm	Initial set of tree	ramped half-and-half method
Selection Algorithm	Method for selecting individuals	lexictour
Crossover	A genetic operator that exchanges subtrees from two parents to form two new children.	rate = 45%
Mutation	A genetic operator that replaces a randomly picked individual's subtree, with a randomly generated one.	rate = 45%
Reproduction	A genetic operator that selects individuals based on fitness and copies them in the next generation.	rate = 10%

where $w = 0.1$. The rationale of the penalty term is that we prefer solution trees that are not very deep and they have a balanced number of variable occurrences with respect to the number of terminal nodes. The value of the weight w was chosen to tune the influence of the penalty term with respect to the PCC term. After several experiments, we found that the value 0.1 was a good tradeoff between solution sizes and results.

The fitness function to be maximized is given by:

$$f(\mathbf{X}) = PCC(\mathbf{X}) - P(\mathbf{X}) \tag{5}$$

In order to apply GP to solve a given problem, several other GP parameters need to be defined. In Table 3 the parameters adopted within our framework are summarized, together with their descriptions.

5 Results and discussions

We have used the GP-framework above described to define our complexity measure. We have adopted the FOOD dataset as training set. The choice of the individual with the best performance might not correspond to the best solution of the problem, due to over-fitting. Over-fitting generally occurs when a model fits overly well the training data, making the model itself not generalizable to new sets of data. To avoid the effect of over-fitting, we have here considered a validation set and we have chosen as best solution of a GP run the individual that presents the best average performance in both training and validation datasets, as already done in the literature.[30,48] As validation set we have chosen the VisCol dataset. Furthermore we have also tested the proposed measure on the VisGray dataset. As the four measures here considered have significantly different ranges, to have a better comprehension of their final combination we have performed a preliminary normalization step. Following a procedure commonly used when applying other machine learning techniques, such as Support Vector Machines (SVM),[49] we have normalized these four measures for the training set mapping the minimum and maximum values to 0 and 1 respectively. During validation and testing, the corresponding measures were rescaled using the extrema of the training set. We have performed 150 GP runs. Averaging over the 150 best solutions, the mean PCC for the training set is PCC= 0.889 while for the validation set is PCC=0.732, with standard deviations of 0.012 and 0.032 respectively.

Among the 150 possible solutions, we have chosen as our complexity measure M_{GP}, the one with the performance on the training set nearest to the obtained mean value. In Fig. 5 the tree representation of this solution is shown. The proposed complexity measure reads as follows:

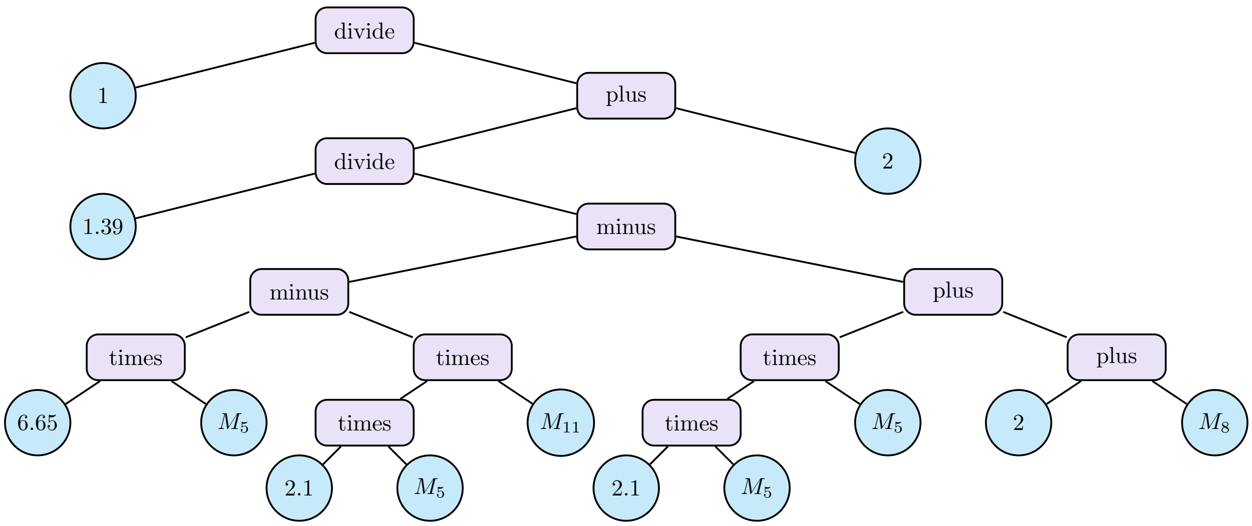

Fig 5 GP tree representing the proposed complexity measure.

Table 4 Performance comparison of our M_{GP} measure against the single measures used in deriving it, and the Visual Clutter measure, on the three datasets.

	$\mathcal{M}_{GP}$	$\mathcal{M}_5$	$\mathcal{M}_8$	$\mathcal{M}_9$	$\mathcal{M}_{11}$	Visual Clutter
training: FOOD	**0.890**	0.862	0.617	0.828	0.426	0.710
validation: VisCol	**0.728**	0.628	0.582	0.695	0.330	0.623
test: VisGray	**0.724**	0.646	0.587	0	0.322	0.529

$$M_{GP} = \left(2 + \frac{1.39}{6.65\, M_5 + M_8 - 2.1\, M_5(M_{11} + M_5) + 2}\right)^{-1} \tag{6}$$

In Table 4 we report the performance in terms of PCC of our M_{GP} measure on each of the three datasets, compared with the performance of the single measures $\mathcal{M}_5$, $\mathcal{M}_8$, $\mathcal{M}_9$, and $\mathcal{M}_{11}$. We also consider the Visual Clutter measure,[50] as a benchmark algorithm. This measure was not specifically designed to predict image complexity but it has shown good correlation in case of real world images[38] and in computer graphic scenes.[51] Our proposal M_{GP} outperforms all the considered single measures, in all cases of training, validation, and test sets.

Trying to better understand the role of the three measures in M_{GP} we rewrite Eq. (6) as follows:

$$M_{GP}(t) = \left(2 + \frac{1.39}{t}\right)^{-1} = \frac{t}{2\,t + 1.39} \tag{7}$$

where

$$t = 6.65\, M_5 + M_8 - 2.1\, M_5(M_{11} + M_5) + 2. \tag{8}$$

However it is not straightforward to understand Eq. (7) as it is. Thus we propose to analyze its Taylor series, centered at a point t_0:

$$M_{GP}(t) = M_{GP}(t_0) + \frac{d}{dt} M_{GP}(t)\bigg|_{t_0} (t - t_0) + \frac{d^2}{dt^2} M_{GP}(t)\bigg|_{t_0} (t - t_0)^2 + o((t - t_0)^3) \tag{9}$$

In Fig. 6 we observe that $M_{GP}(t)$ is well approximated by its Taylor series up to the second order. This helps us in understanding the role of the three single measures starting from Eq. (9), evaluated in $t_0 = 5$:

$$M_{GP}(t) \approx 0.44 + 0.01\,(t - 5) - 0.0047\,(t - 5)^2 \tag{10}$$

Recalling that all the measures were previously normalized in the range $[0\ 1]$, and given t by Eq. (8), we observe that $\mathcal{M}_5$ (Roughness) and $\mathcal{M}_8$ (Number of regions), are the two dominant measures, whose effects sum up through a linear combination. Within this linear combination, their weighting coefficients demonstrate that $\mathcal{M}_5$ gives the highest contribution. The third measure $\mathcal{M}_{11}$ (Memorability) seems to interfere with $\mathcal{M}_5$ introducing a mutual attenuation.

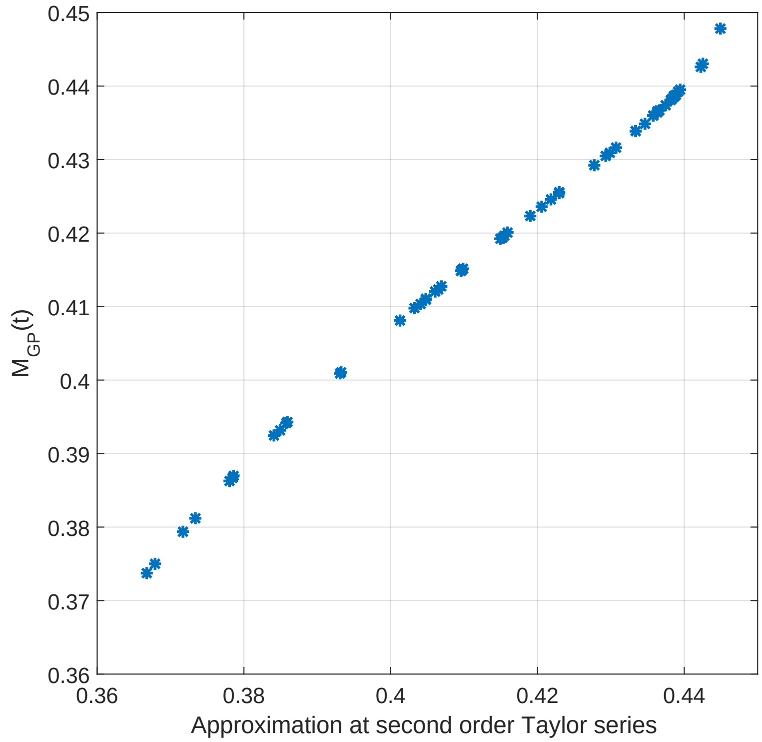

Fig 6 Taylor series up to the second order of M_{GP} evaluated in $t_0 = 5$.

In the proposed M_{GP}, $\mathcal{M}_9$, mainly devoted to evaluate color properties, has not been chosen. However, both M_8 and M_{11} are evaluated on color images and thus color information is included as well. To better understand the influence of each of the four measures considered, we have analyzed all the 150 best solutions. In Table 5 we report the percentage of solutions where each measure appears at least once (solution rate), and the overall frequency of each measure as terminal variable node (variable frequency). From this table it emerges that M_5 is the most frequent measure chosen by the GP. This measure is evaluated on the intensity channel, and was specifically introduced by Tamura to describe texture. Among the remaining three measures, M_{11} (Memorability) is the most frequent one even if, when considered singularly, its performance is the lowest (see Table 4). This suggests that high level measures, that can describe top down mechanisms, have to be also considered to measure complexity perception. The dominant role of a grayscale measure with respect to color measures, agrees with the results obtained from the psycho-physical data. In fact,

Table 5 Solution rates and variable frequencies of the measures derived from the best 150 GP solutions.

	M_5	M_8	M_9	M_{11}
solution rate	96.00%	83.33%	52.00%	91.33%
variable frequency	40.17%	22.03%	13.63%	24.17%

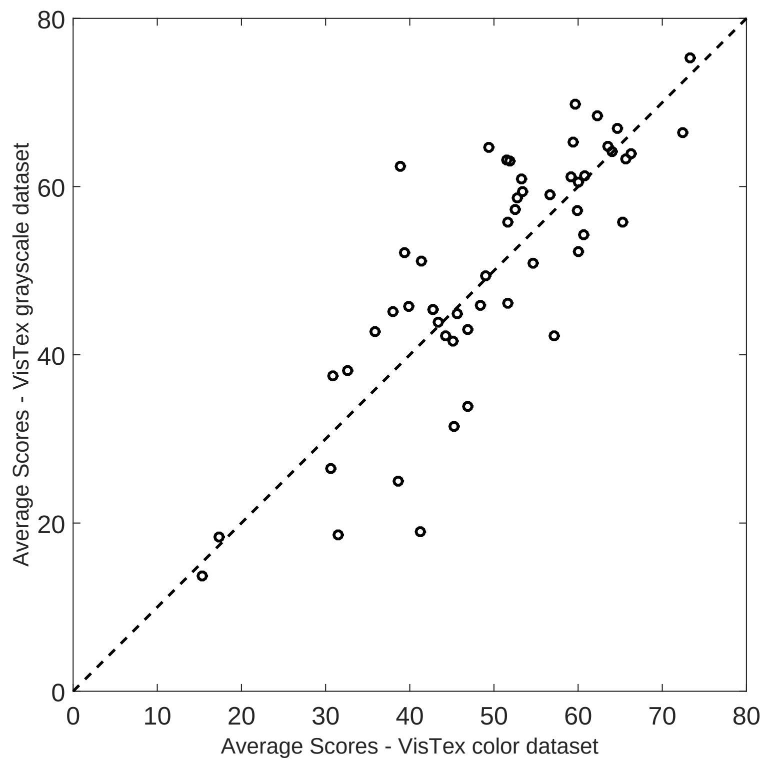

Fig 7 Correlation between subjective score of the VisCol dataset and its grayscale version VisGray.

we found a high correlation (PCC = 0.825) between the subjective scores collected on the VisCol dataset and those collected on the VisGray dataset, as shown in Fig. 7.

6 Conclusions

We have demonstrated that a GP framework is able to design a complexity measure as a nonlinear combination of measures of single image features. This measure outperforms single ones confirming that complexity perception is affected by several visual aspects that mutually interact. We have also found that low level features are suitable to describe complexity perception, but maybe what still lacks to reach higher performance are ad-hoc measures able to better describe top-down aspects, such as familiarity and understandability. The proposed GP framework is generic and allows in the future the combination of other measures that could encode also these properties. To this end, we plan to enlarge the set of image feature measures that can be used in the combination, as well as use other image datasets on which evaluate the complexity. Moreover an extensive analysis of the GP parameters and inclusions of new node functions could be considered to obtain better solutions. One research direction that we plan to pursue is to investigate how image complexity perception depends on the type of semantic content considered (for example texture images versus images representing real world scenes). On this direction we plan to study if a general complexity measure can be derived, in particular considering a dataset that contains images with several different semantic content at the same time.

References

1 A. Rosenfeld, "Visual texture analysis: An overview.," tech. rep., DTIC Document (1975).

2 M. Tuceryan, A. K. Jain, *et al.*, "Texture analysis," *Handbook of pattern recognition and computer vision* **2**, 207–248 (1993).

3 C. Cusano, P. Napoletano, and R. Schettini, "Intensity and color descriptors for texture classification," in *IS&T/SPIE Electronic Imaging*, 866113–866113, International Society for Optics and Photonics (2013).

4 H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions on* **8**(6), 460–473 (1978).

5 R. A. Peters and R. N. Strickland, "Image complexity metrics for automatic target recognizers," in *Automatic Target Recognizer System and Technology Conference*, 1–17 (1990).

6 J. Perkiö and A. Hyvärinen, "Modelling image complexity by independent component analysis, with application to content-based image retrieval," in *Artificial Neural Networks–ICANN 2009*, 704–714, Springer (2009).

7 A. Forsythe, N. Sheehy, and M. Sawey, "Measuring icon complexity: An automated analysis," *Behavior Research Methods, Instruments, & Computers* **35**(2), 334–342 (2003).

8 K. Reinecke, T. Yeh, L. Miratrix, R. Mardiko, Y. Zhao, J. Liu, and K. Z. Gajos, "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*, 2049–2058, ACM (2013).

9 F. Yaghmaee and M. Jamzad, "Estimating watermarking capacity in gray scale images based on image complexity," *EURASIP Journal on Advances in Signal Processing* **2010**, 8 (2010).

10 M. C. Huahui Wu and R. Kinicki, "A study of video motion and scene complexity," in *Tech. Rep. WPI-CS-TR-06-19, Worcester Polytechnic Institute*, (2006).

11 J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao, "Predicting human gaze beyond pixels," *Journal of vision* **14**(1), 28 (2014).

12 Q. Zhao and C. Koch, "Advances in learning visual saliency: From image primitives to semantic contents," in *Neural Computation, Neural Devices, and Neural Prosthesis*, 335–360, Springer (2014).

13 D. C. Donderi, "Visual complexity: a review.," *Psychological Bulletin* **132**(1), 73 (2006).

14 J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*, vol. 1, MIT press (1992).

15 R. Poli, W. B. Langdon, N. F. McPhee, and J. R. Koza, *A field guide to genetic programming*, Lulu. com (2008).

16 C. Cusano, P. Napoletano, and R. Schettini, "Evaluating color texture descriptors under large variations of controlled lighting conditions," *Journal of the Optical Society of America A* **33**, 17–30 (2016).

17 "Mit media lab, vision texture homepage, http://vismod.media.mit.edu/vismod/imagery/visiontexture/."

18 A. N. Kolmogorov, "Three approaches to the quantitative definition ofinformation'," *Problems of information transmission* **1**(1), 1–7 (1965).

19 A. Oliva, M. L. Mack, and M. Shrestha, "Identifying the perceptual dimensions of visual complexity of scenes," in *Proc. 26th Annual Meeting of the Cognitive Science Society*, (2004).

20 J. G. Snodgrass and M. Vanderwart, "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).

21 G. D. Birkhoff, *Collected mathematical papers*, New York (1950).

22 A. R. Rao and G. L. Lohse, "Identifying high level features of texture perception," *CVGIP: Graphical Models and Image Processing* **55**(3), 218–233 (1993).

23 C. Heaps and S. Handel, "Similarity and features of natural textures.," *Journal of Experimental Psychology: Human Perception and Performance* **25**(2), 299 (1999).

24 X. Guo, C. M. Asano, A. Asano, T. Kurita, and L. Li, "Analysis of texture characteristics associated with visual complexity perception," *Optical review* **19**(5), 306–314 (2012).

25 X. Guo, C. M. Asano, A. Asano, and T. Kurita, "Visual complexity perception and texture

image characteristics," in *Biometrics and Kansei Engineering (ICBAKE), 2011 International Conference on*, 260–265, IEEE (2011).

26 G. Ciocca, S. Corchs, and F. Gasparini, "Complexity perception of texture images," in *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, V. Murino, E. Puppo, D. Sona, M. Cristani, and C. Sansone, Eds., *Lecture Notes in Computer Science* **9281**, 119–126, Springer International Publishing (2015).

27 B. Bhanu and Y. Lin, "Object detection in multi-modal images using genetic programming," *Applied Soft Computing* **4**(2), 175–201 (2004).

28 J. Yu and B. Bhanu, "Evolutionary feature synthesis for facial expression recognition," *Pattern Recognition Letters* **27**(11), 1289–1298 (2006).

29 W. Fan, E. A. Fox, P. Pathak, and H. Wu, "The effects of fitness functions on genetic programming-based ranking discovery for web search," *Journal of the American Society for Information Science and Technology* **55**(7), 628–636 (2004).

30 R. d. S. Torres, A. X. Falcão, M. A. Gonçalves, J. P. Papa, B. Zhang, W. Fan, and E. A. Fox, "A genetic programming framework for content-based image retrieval," *Pattern Recognition* **42**(2), 283–292 (2009).

31 M. Iqbal, B. Xue, and M. Zhang, "Reusing extracted knowledge in genetic programming to solve complex texture image classification problems," in *Advances in Knowledge Discovery and Data Mining*, 117–129, Springer (2016).

32 A. Song and V. Ciesielski, "Texture analysis by genetic programming," in *Evolutionary Computation, 2004. CEC2004. Congress on*, **2**, 2092–2099, IEEE (2004).

33 A. Song and V. Ciesielski, “Texture segmentation by genetic programming,” *Evolutionary Computation* **16**(4), 461–481 (2008).

34 W. L. Hays, “Statistics, holt,” *Rinehart and Winston, New York* , 343–383 (1988).

35 D. M. Chandler, “Seven challenges in image quality assessment: past, present, and future research,” *ISRN Signal Processing* **2013** (2013).

36 M. Mack and A. Oliva, “Computational estimation of visual complexity,” in *the 12th Annual Object, Perception, Attention, and Memory Conference*, (2004).

37 H. Yu and S. Winkler, “Image complexity and spatial information,” in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, 12–17, IEEE (2013).

38 G. Ciocca, S. Corchs, F. Gasparini, E. Bricolo, and R. Tebano, “Does color infuence image complexity perception?,” in *Fifth IAPR Computational Color Imaging Workshop (CCIW'15)*, *Lecture Notes in Computer Science* **9016**, 139–148, Springer Berlin / Heidelberg (2015).

39 S. E. Corchs, G. Ciocca, E. Bricolo, and F. Gasparini, “Predicting complexity perception of real world images,” *PloS one* **11**(6), e0157986 (2016).

40 C. Heaps and S. Handel, “Similarity and features of natural textures.,” *Journal of Experimental Psychology: Human Perception and Performance* **25**(2), 299 (1999).

41 P. Isola, J. Xiao, A. Torralba, and A. Oliva, “What makes an image memorable?,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 145–152, IEEE (2011).

42 F. Bianconi, A. Álvarez-Larrán, and A. Fernández, “Discrimination between tumour epithelium and stroma via perception-based features,” *Neurocomputing* **154**, 119–126 (2015).

43 S. Corchs, F. Gasparini, and R. Schettini, “No reference image quality classification for jpeg-distorted images,” *Digital Signal Processing* **30**, 86–100 (2014).

44 D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(5), 603–619 (2002).

45 D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in *Electronic Imaging 2003*, 87–95, International Society for Optics and Photonics (2003).

46 S. Silva and J. Almeida, "Gplab-a genetic programming toolbox for matlab," in *Proceedings of the Nordic MATLAB conference*, 273–278, Citeseer (2003).

47 D. A. Freedman, *Statistical models: theory and practice*, cambridge university press (2009).

48 A. Lacerda, M. Cristo, M. A. Gonçalves, W. Fan, N. Ziviani, and B. Ribeiro-Neto, "Learning to advertise," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 549–556, ACM (2006).

49 C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3), 27 (2011).

50 R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *Journal of Vision* **7**(2), 17 (2007).

51 G. Ramanarayanan, K. Bala, J. A. Ferwerda, and B. Walter, "Dimensionality of visual complexity in computer graphics scenes," in *Electronic Imaging 2008*, 68060E–68060E, International Society for Optics and Photonics (2008).

Gianluigi Ciocca is currently Associate Professor in computer science at DISCo (Department of Informatics, Systems and Communication) of the University of Milano-Bicocca. His research focused on the development of systems for the management of image and video datasets and the development of new methodologies and algorithms for automatic indexing. His current research interests focus on image and video analysis and abstraction. He is member of NeuroMi, Milan Center for Neuroscience.

Silvia Corchs is a Senior Research Associate at DISCo (Department of Informatics, Systems and Communication) of the University of Milano-Bicocca. She has PhD degrees in Physics and in Computer Science. Her research focused on atomic physics and the study of computational models for the attentional mechanism. At present, the current research interests focus on computational neuroscience and image processing. She is member of NeuroMi, Milan Center for Neuroscience.

Francesca Gasparini took her degree in Nuclear Engineering at the Polytechnic of Milan in 1997 and her Ph.D in Science and Technology in Nuclear Power Plants at the Polytechnic of Milan in 2000. She is currently assistant professor at DISCo (Department of Informatics, Systems and Communication) of the University of Milano-Bicocca. Her main research topics are related to image and signal processing and multimedia applications. She is member of NeuroMi, Milan Center for Neuroscience.

List of Figures

1 GP tree representing the program: $5 \times X_1 + sin(X_2) + X_3$

2 Thumbnails of the texture images chosen to sample 58 classes of the RawFooT dataset, row-wise ordered from the less complex to the most complex, according to the Average Score.

3 Thumbnails of the texture images chosen to sample each of the 54 classes in the VisTex dataset, row-wise ordered from the less complex to the most complex), according to the Average Score.

4 Thumbnails of the texture images chosen to sample each of the 54 classes in the VisTex dataset in their grayscale version, row-wise ordered from the less complex to the most complex, according to the Average Score.

5 GP tree representing the proposed complexity measure.

6 Taylor series up to the second order of M_{GP} evaluated in $t_0 = 5$.

7 Correlation between subjective score of the VisCol dataset and its grayscale version VisGray.