IET Computer Vision

Research Article

ISSN 1751-9632
Received on 7th June 2014
Revised on 6th March 2015
Accepted on 2nd April 2015
doi: 10.1049/iet-cvi.2014.0141
www.ietdl.org

Measuring meaningful information in images: algorithmic specified complexity

Winston Ewert[1], *William A. Dembski*[1], *Robert J. Marks II*[2] ✉

[1]*Evolutionary Informatics Laboratory, McGregor, TX 76657, USA*
[2]*Department of Electrical and Computer Engineering, Baylor University, Waco, TX 76798-7356, USA*
✉ *E-mail: RJMarksII@gmail.com*

Abstract: Both Shannon and Kolmogorov–Chaitin–Solomonoff (KCS) information models fail to measure meaningful information in images. Pictures of a cow and correlated noise can both have the same Shannon and KCS information, but only the image of the cow has meaning. The application of ‘algorithmic specified complexity’ (ASC) to the problem of distinguishing random images, simple images and content-filled images is explored. ASC is a model for measuring meaning using conditional KCS complexity. The ASC of various images given a context of a library of related images is calculated. The ‘portable network graphic’ (PNG) file format’s compression is used to account for typical redundancies found in images. Images which containing content can thereby be distinguished from those containing simply redundancies, meaningless or random noise.

1 Introduction

Humans can readily distinguish meaning in images. However, what is our theoretical basis for doing so? If we look at a picture of a sunset, we readily identify it as not being a random assortment of pixels, but why? Generating an image such as a sunset by randomly choosing pixels is astronomically improbable. However this is also true of any given image – even one of pure noise. The image of a sunset has more meaningful information than that of an image of random noise. A bit count alone does not measure meaning. The number of bit can be the same for both images.

Although the term ‘information’ is commonly used, its precise definition and nature can be illusive. If we shared a digital versatile disc (DVD), is information being destroyed? What if there are other copies of the DVD? Is information being created when we snap a picture of Niagara Falls? Would a generic picture of Niagara Falls on a post card contain less information than the first published image of a bona fide extraterrestrial being? These questions cannot be answered properly with a direct ‘yes’ or ‘no.’ An elaboration on the specific definition of ‘information’ being used is first required. Shannon recognised his formulation of information could not be used in all contexts [1, 2].

> “It seems to me that we all define ‘information’ as we choose; and, depending on what field we are working in, we will choose different definitions. My own model of information theory... was framed precisely to work with the problem of communication.”

As a result, different formulations of different information measures have been proposed to fit various problems. Shannon information [Thermodynamic entropy motivated Shannon’s naming of (Shannon) entropy [3]. Thermodynamic entropy is often viewed through the lens of Shannon information [4]. See, for example, Bekenstein [5].] [4, 6, 7] and Kolmogorov–Chaitin–Solomonoff (KCS) complexity [4, 8–15] have served as the foundation in these proposed model variations [16–21].

For an image to be meaningfully distinguishable, it must relate to some external independent pattern or specification. The image of the sunset is meaningful because the viewer experientially relates it to other sunsets in their experience. Any image containing content rather than random noise fits some contextual pattern. Naturally, any image looks like itself, but the requirement is that the pattern must be independent of the observation and therefore the image cannot be self-referential in establishing meaning. External context is required.

If an object is both improbable and specified, we say that it exhibits ‘specified complexity’ [22–25]. A page of kanji characters, for example, will have little specified complexity to someone who cannot read Japanese.

A striking example is the image in Fig. 1. On first viewing, the image seems to have no specified complexity. During prolonged viewing, the mind scans its library of context until the meaning of the image becomes clear.

1.1 KCS complexity

KCS complexity is defined as the length of the shortest programme required to reproduce a result, in this case the pixels in an image. KCS complexity is formally defined as the length of the shortest computer programme, p, in the set of all programmes, P, that produces a specified output X using a universal Turing machine, U

$$K(X) = \min_{U(p,)=X|p\in P} |p|$$

Such programmes are said to be ‘elite’ [14]. ‘Conditional KCS complexity’ [9, 10] allows programmes to have input, C, which is not considered a part of the elite programme

$$K(X|C) = \min_{U(p,C)=X|p\in P} |p|$$

C is the ‘context’.

The more the image can be described in terms of a pattern, the more compressible it is, and the more specified. For example, a black square is entirely described by a simple pattern, and a very short computer programme suffices to recreate it. As a result, we conclude that it is highly specified. In contrast, an image of randomly selected pixels cannot be compressed much if at all, and thus we conclude that the image is not specified at all. Images with content such as sunsets take more space to describe than the black square, but are more specified than random noise. Redundancy in some images is evidenced by the ability to approximately restore groups of missing pixels from those remaining [27, 28].

Fig. 1 *Image used to demonstrate the difference between eyesight and vision*

Initially, this image appears to be only random splotches of grey. After prolonged viewing, however, the mind finds context by which to interpret the image.
Once the context is established and the image seen, subsequent viewing will immediately revert to the contextual interpretation of the image. The object in the picture is a cow. The head of the cow is staring straight out at you from the centre of the photograph, its two black ears framing the white face. The picture is widely used by the Optometric Extension Program Foundation to demonstrate the difference between eyesight and vision [26]

An image of uniform random noise defies compression. Other images with stochastic components may be compressible. For example, a large square with uniform grey level on a black background is described by a distribution with probability mass at only two locations and is consequently highly compressive. Small amounts of noise about this grey level will also be compressible, but to a lesser extent. It would seem problematic to classify such a simple image with the images of sunsets or other content. To account for this, we obliged to model a stochastic process which can produce such simple images. Which images might be considered simple depends on the stochastic process being modelled.

1.2 Algorithmic specified complexity

Given a particular stochastic process, we would like to be able to measure how well a given image is explained by a particular given stochastic process. The goal is to separate those images which look like they were produced by the stochastic process from those which were not. Towards this end we define 'algorithmic specified complexity' (ASC) [22, 23, 25] as

$$\mathrm{ASC}(X,\ C,\ P) = I(X) - K(X|C) \tag{1}$$

where X is the object or event or under consideration, C is the context, given information which can be used to describe the object, $P(X)$ is the probability of X under the given stochastic model, $I(x) = -\log_2 P(X)$ is the corresponding self-information and $K(X|C)$ is the conditional KCS complexity of X given context C.

By taking into account the conditional KCS complexity and the probability assigned by the stochastic process, the ASC measures the degree to which an image fits the hypothesised stochastic process. Given high ASC, we have reason to believe that the image is unlikely to be produced by that process. In fact, the occurrence of images with high ASC is rare. Specifically [23]

$$\Pr[\mathrm{ASC}(X,\ C,\ P) \geq \alpha] \leq 2^{-\alpha} \tag{2}$$

Thus, bounding the probability of obtaining high ASC images when sampled according to a given distribution. For example, since $2^{30} \simeq 10^9$, we have about one in a billion chance of obtaining 30 bits of ASC. A large ASC is strong indication that an image was not produced by the proposed stochastic process.

For the ASC to be small, the conditional KCS complexity must be small in comparison to the self-information term. However, both of these quantities must be taken into account before announcing the degree of meaning in an object. The conditional KCS might be small because the unconditional KCS is small. Therefore the ASC cannot be ascertained by inspection of the conditional KCS complexity alone. The self-information term is mandatory for indirectly assessing whether the conditional KCS complexity is small because of rich context or because the original unconditional KCS complexity is small.

Since KCS complexity is incomputable, ASC is incomputable [4, 14, 29]. However, the true KCS complexity is always equal to or less than any known estimate of it. We will refer to a known estimate as the 'observed ASC' (OASC). We know that

$$\mathrm{ASC}(X,\ C,\ P) \geq \mathrm{OASC}(X,\ C,\ P) \tag{3}$$

Thus $\mathrm{OASC}(X, C, P) = \mathrm{ASC}(X, C, P) - k$ for some $k \geq 0$ and

$$\begin{aligned}\Pr[\mathrm{OASC}(X,\ C,\ P) \geq \alpha] &= \Pr[\mathrm{ASC}(X,\ C,\ P) - k \geq \alpha] \\ &= \Pr[\mathrm{ASC}(X,\ C,\ P) \geq \alpha + k] \\ &\leq 2^{-\alpha-k} \\ &\leq 2^{-\alpha}\end{aligned} \tag{4}$$

OASC therefore obeys the same bound as ASC.

ASC is defined based on conditional KCS complexity. The context enables compression to take advantage of known information. A picture of a house defies explanation by a simple stochastic process alone. If we take the context to be a library of known images, then the similarity should allow us to describe the new image by making use of details from the library images. Without the context, images with simple patterns such as simple shapes or fractals [Interestingly, fractal patterns are well known to be highly compressible [4] and therefore have an extremely low KCS complexity. Their KCS complexity is low with or without context. The ACS of a fractal image will be high if an ill-informed stochastic model generates a large self-information. If, on the other hand, the stochastic model includes fractal structures, the corresponding ACS will be low.] could be deemed compressible, but it is difficult to see that an image of a house alone would be compressible. Including context lets us take into account prior experience and area of knowledge.

Note that the ASC measure is not simply labelling a picture as belonging to a category such as 'houses.' ASC, rather, measures the difficulty of generating the digital picture of the house exactly to the pixel level.

A solid black square may be assigned a high probability by a reasonable stochastic process. It is very compressible and thus specified, but does not have a level of ASC because of its low complexity. A random image will be assigned a low probability by a stochastic process, but it is not compressible and therefore not specified. As a result, it will not have a high value of ASC either. A sunset will be given a low probability by a stochastic process (excluding those designed to produce images of sunsets). It is also specified because it can be described by a shorter computer programme. Consequently, the ASC of the sunset image will be high. The ASC allows us to distinguish between these various categories of images.

By using a library of images in a number of scenarios, we demonstrate ASC's ability to distinguish images with contextual meaning from those with without. ASC is illustrated for noise, algorithmic transformations and different camera shots of the same object.

1.3 Background

1. History: The idea of ASC model was first presented by Dembski [22]. The topic was developed and illustrated with a number of examples [23, 25]. Durston *et al.*'s 'functional information' model [19] was shown to be a special case of ASC. Application to intricate artificial life-like patterns designed around Conway's 'Game of Life' show that the ASC can be useful in more complex environments [24]. Additional history concerning the development of ASC can be found in our previous work on the subject [23, 24].
2. Distinction: ASC differs from conventional signal and image detection [30–35] including matched filter correlation identification of the index of one of a number of library images [36–38]. Alternately, KCS complexity asks for the minimum information requirements to reproduce an image losslessly (i.e. exactly) – pixel by pixel.
3. The meaning of meaning: KCS complexity has been used to measure meaning in other ways. Kolmogorov 'sufficient statistics' [4, 29] can be used in a two part procedure to identify the algorithmically random component of X. The remaining non-random structure can then be said to have 'meaning' [39]. The term 'meaning' here refers to the internal structure of the object under consideration and does not consider the context available to the observer as is done in ASC.
4. Mixing Shannon and KCS information models: The ASC model in (1) combines a probabilistic Shannon model with the KCS model of information. Although the KCS and Shannon models are often thought of as distinct, they often yield commensurate results. The expected value of the KCS complexity of a random string of bits, for example, is close to the corresponding Shannon entropy [4]. The KCS complexity X is approximately equal to the Shannon self-information corresponding to the 'universal probability' of randomly choosing a computer programme to generate X [4, 29]. The difference of the KCS complexity from the Shannon self-information determined by universal probability is dubbed the 'randomness deficiency' [29].
5. KCS complexity applied to images: On the basis of the notion of information distance [40], KCS complexity has been proposed as a tool to compute image similarity [41, 42]. The method uses the similarity between two binary sequences (or anything mapped to binary sequences) using conditional KCS complexity. Specifically, if two images are similar, there should be a set of algorithmic transformations to convert one image into the other such that less space is required to describe the transformations than to simply encode the image directly. Others have worked on the problem of compressing similar images [43, 44]. The idea is that we should be able to take advantage of image similarities to compress them better. The compressibility of similar images is also fundamental for the work considered here. Without it, using a library of images to compress related images would not be possible as is discussed in Section 2.6.2.
6. Relation of ASC to mutual information: ASC models a methodology whereby humans can assess meaning from sensory inputs and their experience. According to Tononi [21], consciousness can be measured in terms of integrated information denoted by Φ. Gregory Chaitin (the C in KCS) recently opined [45]:

> "I suspect Φ has something to do with what in algorithmic information theory is called mutual information... which is the extent to which X and Y are simpler when seen together than when seen separately."

The ASC measure in (1) bears a resemblance to Shannon mutual information [7, 4] as a function of Shannon entropy and conditional entropy

$$I(X;\ Y) = H(Y) - H(Y|X)$$

Shannon mutual information is a measure of the dependence of two random variables X and Y. The maximum of the mutual information is the channel capacity which determines the maximum rate communication can occur over the channel without error. The KCS version of mutual information is [29]

$$I_K(X;\ Y) = K(Y) - K(Y|X)$$

In the same spirit, the ASC measure in (1) can be thought of as measuring the resonance between known context and observation with respect to an interpretive model.

2 Measuring meaning in images

We now show how ASC can be applied to measuring meaning in images.

2.1 Image library

Fig. 2 shows three pictures of famous scientists which make up the library of images for our context in this example. For contrast, see Fig. 3 which shows a solid square and an image of random noise. These two images are not in the library. The square is very compressible because of its single solid colour, whereas the random image is not. Random noise does generally not compress well.

In the simplest case, we want to compress an image exactly identical to one in the library. We can easily describe such an

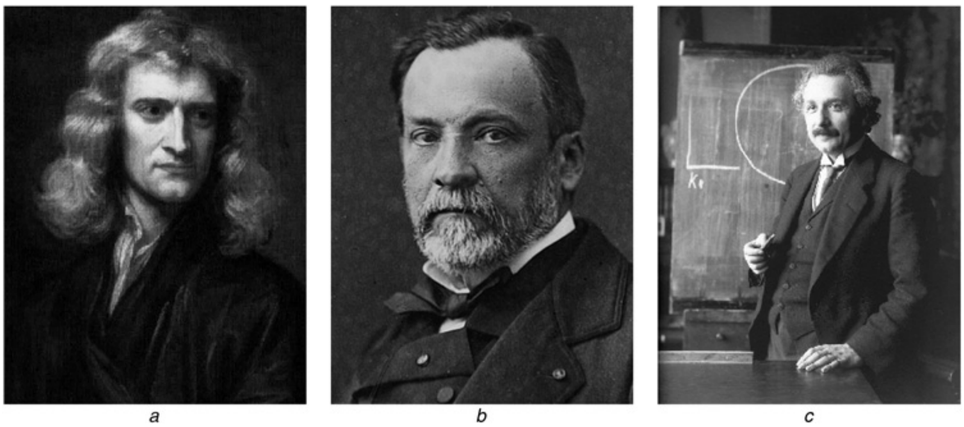
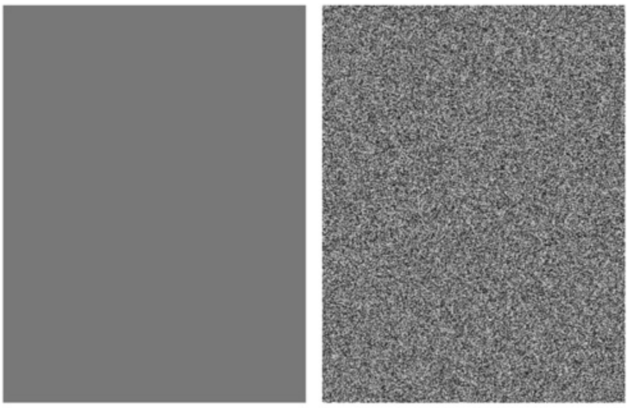


Fig. 2 *Images of scientists*

a Newton
b Pasteur
c Einstein

IET Comput. Vis., 2015, Vol. 9, Iss. 6, pp. 884–894



a b

Fig. 3 *Comparison images not included in the library*

a Solid grey square
b Random image

image merely by its index in our small library. Thus [We adopt the commonly used notation $A \overset{+}{<} B$ to mean $A < B + c$ where c is a constant. (See e.g. Bennett *et al.* [40] and Grünwald *et al.* [8].) For example, KCS complexity differs from Turing machine to Turing machine, but is equal up to a constant allowing translation of one Turing machine language into the other [4, 29]. The length of the translating programme is independent of the object being compressed. The c will vary from computer to computer and description format to description format. Similarly $A \overset{-}{<} B$ means $A < B - c$.]

$$K(X|C) \overset{+}{<} \lceil \log 3 \rceil = 2\,\text{bits} \quad (5)$$

The images are 284 × 373 pixels in grey scale, with $2^8 = 256$ levels of grey. The raw grey-scale image encoded directly would require 8 × 284 × 373 = 847 456 bits. Initially, we will postulate the images were generated by randomly choosing the grey scale for each pixel uniformly across all 256 possible values. This would mean that every possible grey-scale image has an equal probability

$$\Pr[X] = 2^{-847\,456} \quad (6)$$

where X is the random variable constituting the image. The Shannon self-information of an image from this population is then

$$I(X) = -\log_2 \Pr[X] = -\log_2 2^{-847\,456} = 847\,456\,\text{bits} \quad (7)$$

Using the formula for ASC in (1) and the three images as context, we obtain for any one of the library images

$$\text{ASC}(X,\ C,\ P) \geq \text{OASC}(X,\ C,\ P) \doteq 847\,456 - 2 = 847\,454\,\text{bits}$$

The rich context provided by the three image library results in each of the scientist images having significant meaning. Recall that $\Pr[\text{ASC} > 847\,454] \leq 2^{-847\,454}$ which renders the probability of generating these images through such a stochastic process as absurdly improbable.

How does the process fare for a simple pattern such as a library of equally sized solid squares differing only in grey scale? The square can be described by its shade of grey which requires 8 bits for 256 grey levels. Using this context, the complete description of a solid square image is

$$K(X|C) \overset{+}{\leq} 8\,\text{bits} \quad (8)$$

Thus, the OASC for the solid square of the same size as the scientists' pictures would be OASC$\doteq 847\,456 - 8 = 847\,448$ bits. The square is only slightly less likely to be produced by the stochastic process than the detailed images of the scientists. This is because randomly choosing all pixels with the same grey level using the uniformly distributed stochastic model is extremely unlikely. The stochastic process we are using does not assign higher probability to simple patterns.

However, we now define another stochastic process which does so.

2.2 Self-information based on portable network graphic (PNG) compression

Lossless compression algorithms can be used to estimate ASC. Commonly used lossless compression algorithms are based on Lempel–Ziv compression [46, 47] later improved by Welch to Lempel–Ziv–Welch (LZW) compression [4, 47, 48]. The algorithm is used in PKZIP [49], DEFLATE [50] and WinZip [51]. 'Graphics interchange format (GIF)' image compression is similarly dependent on LZW compression. The limited abilities of GIF compression has been replaced by the 'PNG' compression [52, 53] which is similarly based on the LZW algorithm.

We will adopt an approximation of complexity based on length of PNG files. The widely used PNG format is designed to take advantage of certain redundancies present in images to produce better lossless compression. Thus, the modelled stochastic process will produce images containing these sorts of redundancies. Redundancies such as found in the library of solid squares will not generate large values of self-information using PNG's and therefore do not provide the basis for a high ASC.

The first 8 B of a PNG image file are always the same, so we have excluded these from the length calculation. We assume that the probability of an image is thus

$$\Pr[X] = 2^{-\ell(X)-8} \quad (9)$$

where $\ell(X)$ is the length in bits of the PNG file required to produce the image. Naturally, this gives a self-information value of $I(X) = \ell(X) - 8$.

Table 1 shows the complexity and ASC for various images under the two different stochastic models. The pictures of the scientists all compress to similar lengths in PNG and are thus deemed similarly complex. The random image is significantly more complex, whereas the solid square is much less complex. Using the PNG complexity, the square image with its redundant pixel values has two orders of magnitude less ASC than the other images. The square image is much better explained than any of the library images. It still has a large amount of ASC due, in part, to the high unlikeliness of creating a solid image by randomly generating PNG files.

An initially somewhat surprising result is the quantity of ASC found in the random image when using the PNG complexity measure. As might be expected, under a uniform distribution over the 256 possible grey levels, the complexity and specification cancel each other out leaving absolutely no indication of specified complexity. However, the PNG-based stochastic model assigns lower probabilities to images lacking any sort of redundancy. The absence of redundancy means that the image does not fit the modelling stochastic process.

Table 1 Details on the various images

Image	Complexity (uniform)	Complexity (PNG)	KC	OASC (uniform)	OASC (PNG)
Newton	847 456	520 224	2	847 454	520 222
Pasteur	847 456	543 000	2	847 454	542 998
Einstein	847 456	513 064	2	847 454	513 062
square	847 456	6224	8	847 448	6216
random	847 456	849 008	847 456	0	1552

Constant *c* is omitted. KC denotes the conditional KCS complexity

2.3 Noise

Not all images will be identical to those in the library. For a simple case consider a noisy copy of an image. The image is the same as the library version, except that noise has been added to it. To compress the image, we need to specify both the image in the library as well as the noise.

(a) For the three images of scientists

$$K(X|C) \overset{+}{\leq} \lceil \log_2 3 \rceil + \mathrm{p}H(N) \tag{10}$$

where p is the number of pixels and $H(N)$ is the Shannon entropy of the noise N [54]. Note that only the entropy of the random variable affects the description length. If we ignore bit levels saturating at 0 and 255, the mean of the variable can be shifted without forcing the image to use any additional space.
(b) The square image cannot be described as similar to the one in the library, but it can be described as its base colour with the noise

$$K(X|C) \overset{+}{\leq} 8 + \mathrm{p}H(N) \tag{11}$$

(c) More generally, adding noise to a random image produces another random image leaving us with no way of compressing it. Thus

$$K(X|C) \overset{+}{\leq} 8p \tag{12}$$

We can now view the ASC as a function of noise for the running example. Fig. 4, for example, shows the picture of Pasteur as increasing levels of noise are added. We add uniform random noise to each pixel. Saturated pixels are shown as either black or white. Fig. 5 shows the plot of the varying images as levels of noise are increased. At 0% noise, the image is exactly identical to the one in the library. At 100% noise, the image is indistinguishable from random noise. The ASC of Einstein and Newton images follow similar curves. There is initially a great deal of ASC, but this decreases as the noise is increased. Interestingly, the square has an initial increase in ASC as noise is added. This is because the PNG file format works very well to compress a solid square, but does a relatively poor job of compressing that square with just a small amount of noise.

There is a relatively flat period between 20 and 60%. This is caused by a closely matched increase in the PNG length of the images and the KCS complexity of those images. The noise increases both the complexity of the image as well as decreasing the specification. These two changes cancel out leaving a slow change. All of the methods tend towards zero ASC as the noise reaches 100%.

As expected, the curve for the random image in Fig. 5 is flat and exhibits very low amounts of ASC.

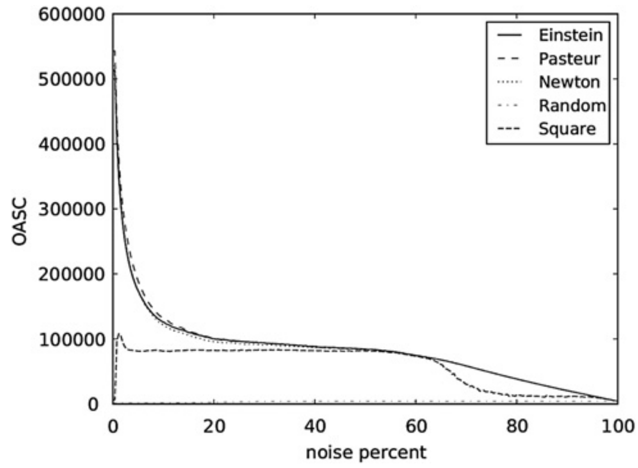


Fig. 5 *ASC for varying levels of noise*

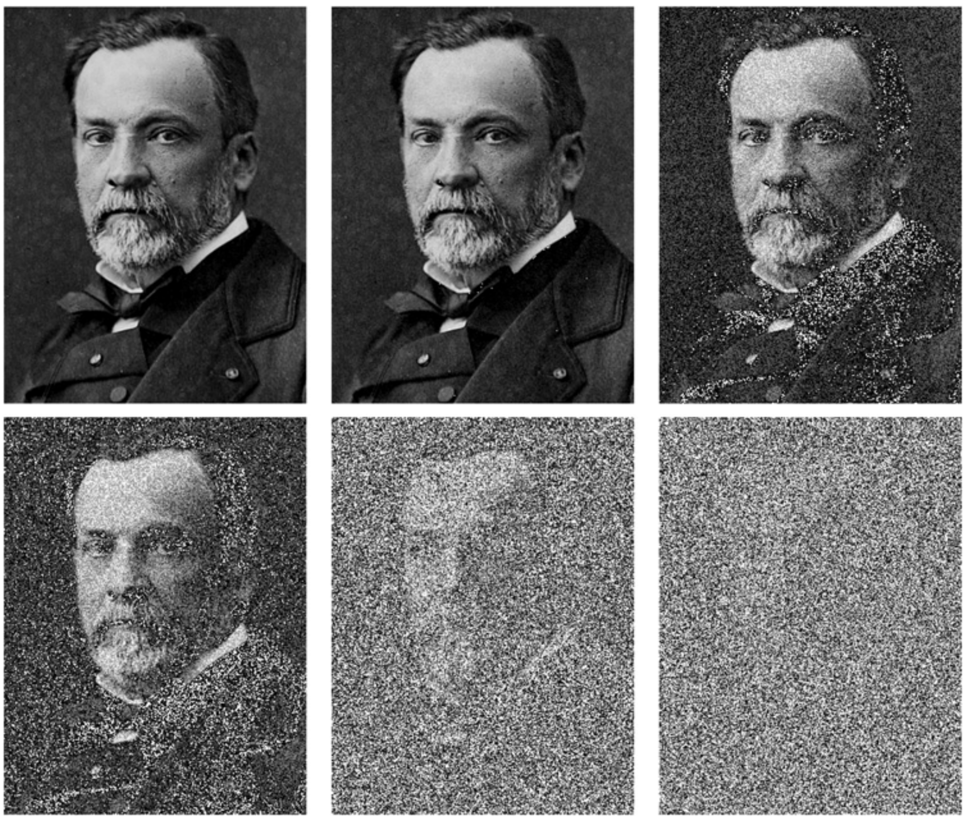


Fig. 4 *Picture of Louis Pasteur with increasing levels of added noise*

IET Comput. Vis., 2015, Vol. 9, Iss. 6, pp. 884–894

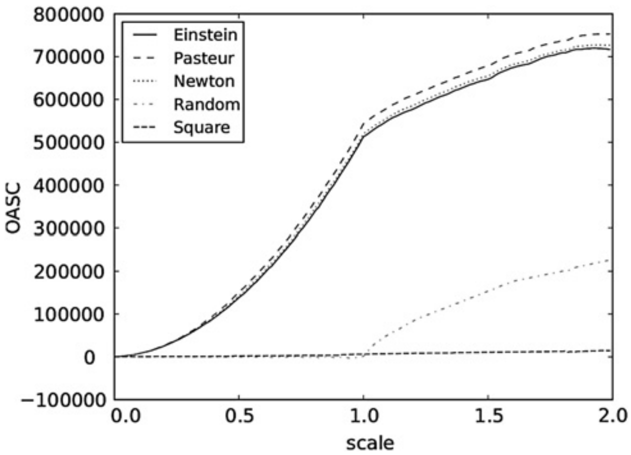


Fig. 6 *OASC for different resizings*

2.4 Scaling

Another possible perturbation of library images on images is scaling. In this case, we should be able to resize the image from the library to match the one we are compressing. As long as the image has been resized in an algorithmic way, we can describe the image by specifying the value from the library along with the scaling factor. There are many different possible scaling algorithms, but they will all simply result in a different constant c for the programme length. We will represent the scaling factor as $(x/1000)$ and allow scaling factors from 0 (the image is resized to an image of zero width and height) to almost 2 (the image is doubled in size). This corresponds to 2000 different scalings.

(a) We can encode each scientist's image as the index from the library along with the scaling factor

$$K(X|C) \overset{+}{\leq} \lceil \log_2 3 \rceil + \lceil \log_2 2000 \rceil \tag{13}$$

(b) The solid square has to be described as the shade of grey and the scaling factor

$$K(X|C) \overset{+}{\leq} 8 + \lceil \log_2 2000 \rceil \tag{14}$$

(c) Finally, for the random image, scaling up can be described as the original random image and the scaling factor

$$K(X|C) \overset{+}{\leq} 8p + \lceil \log_2 2000 \rceil \tag{15}$$

where p is the number of pixels in the pre-scaled image. However, KCS complexity is defined as the shortest programme that produces the result and this is not the most efficient method to describe a scaled down random image. Rather we can encode the image directly

$$K(X|C) \leq 8\,s \tag{16}$$

where s is the number of pixels in the scaled image. Note that when $s = p$ both methods will be approximately equal in length.

Fig. 6 shows the ASC for the images and varying resizes. For the scientists, the OASC increases as the scale does. It increases quickly for scales below one, whereas it increases slowly for scales above 1. This is because scaling up the original images introduces redundancy into the images which PNG compresses. Thus, the complexity increases slowly. Scaling down the image loses information, thus exhibiting a rapid decrease in OASC. This is evident in Fig. 7 where scaled down versions of Einstein are shown magnified. On the right, for example, the details of the vest buttons and of the pencil Einstein is holding have been obliterated. Random noise slows the OASC increase after passing the 1.0 point as well. Although the base image is random, redundancy is introduced by the scaling process.

2.5 Repeated element

Figs. 8*a* and *b* show two images which both share a stick man figure. Otherwise the images are random noise. Using the image in Fig. 8*a* as our context, we will attempt to compress the image on the right. The second image can be described as the stick figure from the first image together with the difference encoded as an image. The difference is shown in Fig. 8*c*. Note that the noise in the bounding box of the stick man in Fig. 8*c* is calculated such that adding it to the noise around the stick figure in the library image will produce the noise from the target image. Table 2 shows the number of bits required to describe the images by PNG. To actually describe the image then requires specifying the bounding box of the stick man in the original image (four coordinates) as well as the target in the current image (two coordinates). Since the images are 400×400 pixels, this requires

$$6 \log_2 400 \simeq 52\,\text{bits} \tag{17}$$

Thus

$$K(X|C) \overset{+}{\leq} 52 + \ell \tag{18}$$

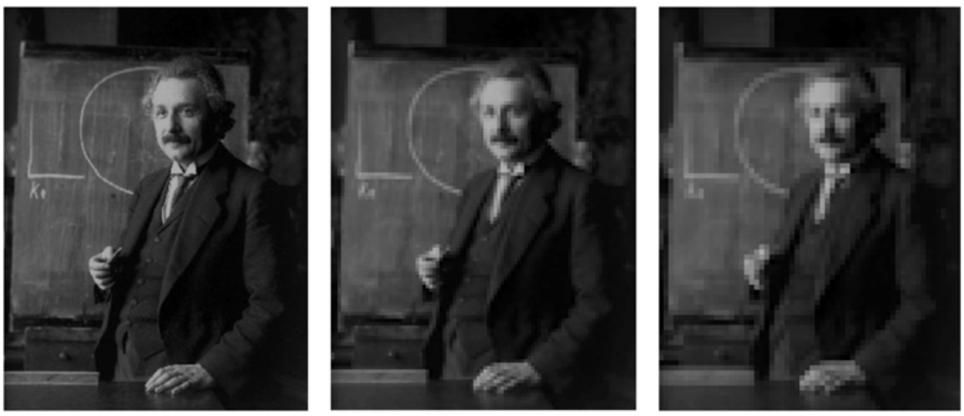


Fig. 7 *Magnified scales of Einstein*

Images are scaled using bicubic interpolation [54]
Left: original, middle: (1/4) scale and right: (1/6) scale

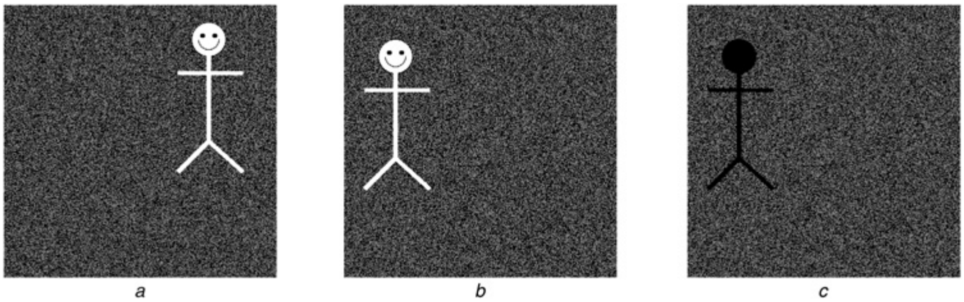


Fig. 8 *Stick men on a sea of noise*

a Context stick man
b Stick man image
c Difference image

where ℓ is the length of the PNG compression of the difference image. In this case, $\ell = 211\,576$ so $K(X|C) \overset{+}{\leq} 211\,628$ and $\text{ASC} \geq 216\,496 - 211\,628 = 4868$ bits. The object being in a different location and the random background noise did not prevent ASC from being observed. The target image contains information by virtue of containing the same stick figure as the original image.

2.6 Photographs

2.6.1 Offset and difference OASC:

Two photographs taken of the same object will differ slightly in all sorts of ways. For example,

Table 2 PNG complexity length for the various man images

Name	PNG complexity
context	216 568
image	216 912
difference	211 712

the picture may be shifted and the noise different. Fig. 9 shows a collection of images [55]. Each image is representative of a collection of photos taken of the same object from slightly varying positions. These images can be aligned by shifting the image by an offset. We take these representative images as our context, and attempt to compress other images in the collection. We do this by recording the needed offset as well as a difference image; samples of which are shown in Fig. 10. Each image can be described as

$$K(X|C) \overset{+}{\leq} \log_2|L| + \log_2 w + \log_2 h + \ell \qquad (19)$$

where L is the set of images in the library, w and h are the height of the image and ℓ is the PNG length of the difference image. The $\log_2|L|$ term is to determine which image from the library should be used. The w and h are present to specify the offset between the library image and the image under inspection.

Figs. 11–16 show scatter plots of the OASC. Each point is a single image's ASC using the context of the images shown in Fig. 9. The x-axis is the Manhattan distance of the shift required to line up the two images. For most of the collections, the ASC moves towards

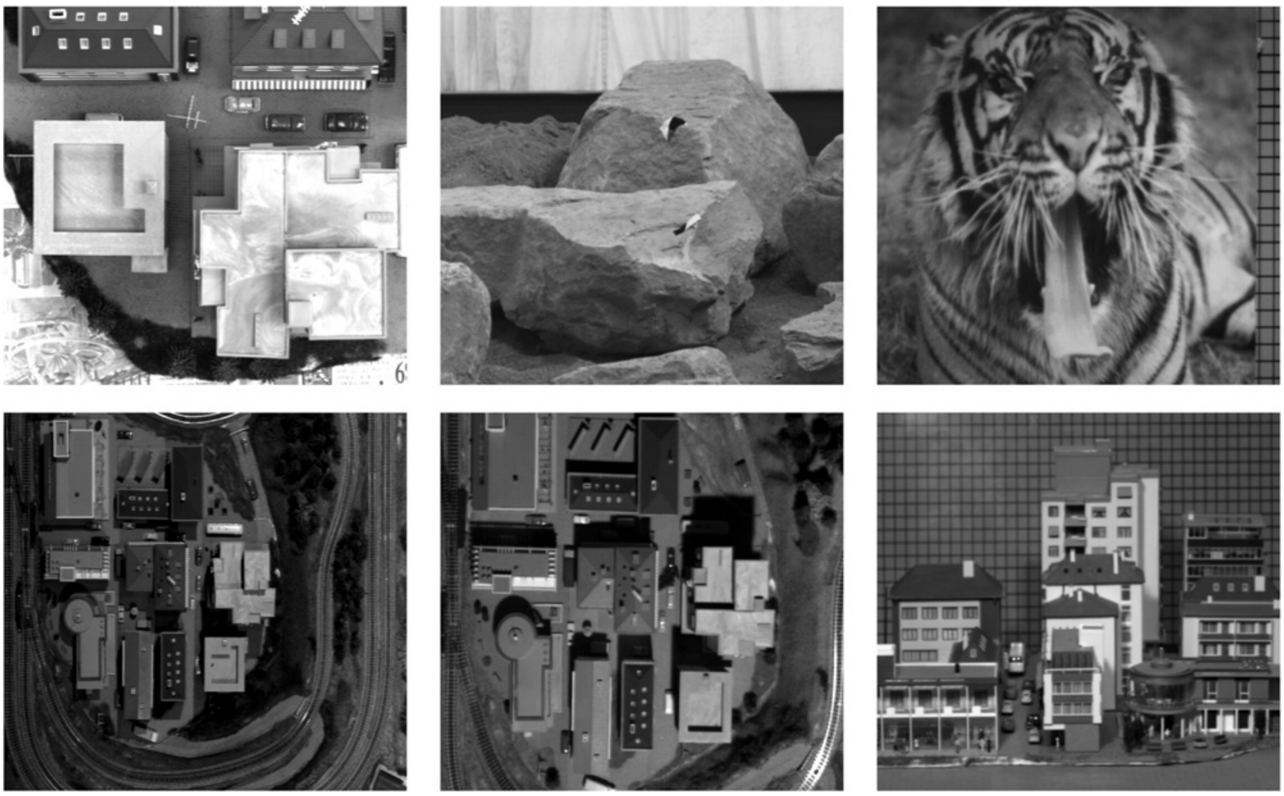


Fig. 9 *Collection of images*

IET Comput. Vis., 2015, Vol. 9, Iss. 6, pp. 884–894

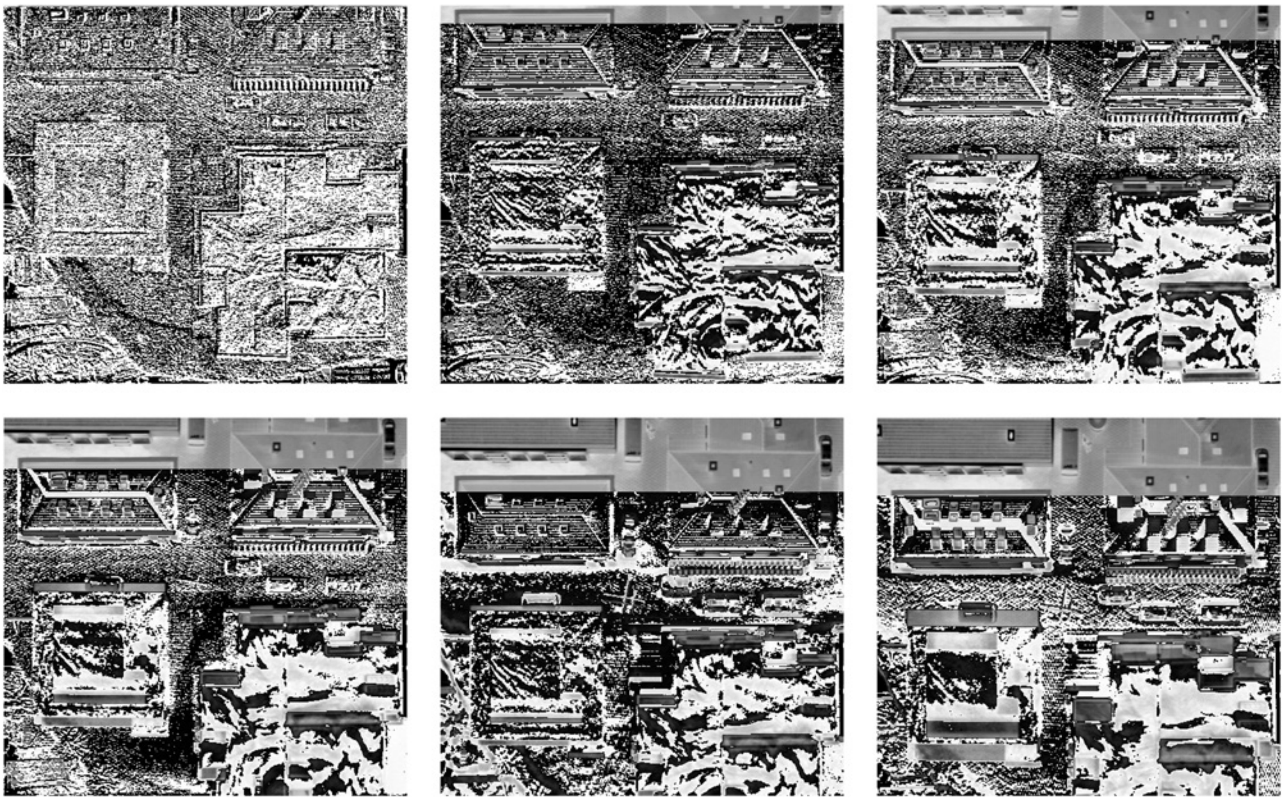


Fig. 10 *Aerial city shot difference images*

zero as the required shift increases. An exception is the tiger images in Fig. 14 which maintain most of their ASC value. Fig. 12 has an outlier where the difference image compressed poorly, but the overall trend remains. This is because the tiger image is a photograph of a photograph and thus lacks three-dimensional effects. However, images with small shifts contain significant amounts of ASC. This means that we can conclude that the other images are not simply random noise. They share too much similarity with the random image to be generated by a stochastic process, even one that introduces redundancies into images.

2.6.2 ASC from measuring compression file sizes only:

Compression algorithms used in Section 2.2 to evaluate the self-information term in ASC can also be used to estimate KCS complexity [56–58]. The size of the compressed object X is an upper bound for $K(X)$. We will call this estimate $K_O(X)$.

To illustrate the potential use of compression in evaluating OASC, consider again the images of Newton and Pasteur in Fig. 2. Both images are scaled to 300×400 pixels. Assuming a byte per pixel and a random stochastic model for image generation, both therefore have a self-information of

$$I(N) = I(P) = 300 \times 400 \times 8 = 960\,000\,\text{bits} = 120\,\text{kB} \quad (20)$$

where we have used P for Pasteur and N for Newton. The PNG file sizes for the two images are

$$K_O(N) = 74\,\text{kB} \quad \text{and} \quad K_O(P) = 76\,\text{kB} \quad (21)$$

Consider, then, placing identical images of Newton side-by-side forming a 600×400 image. The number of pixels has doubled. We expect that $K(X, X) \stackrel{+}{=} K(X)$. The PNG compression captures the redundancy since the size of the side-by-side images is $K_O(N, N) = 77$ kB. This is just a tad more than 74 kB in (21). We can do

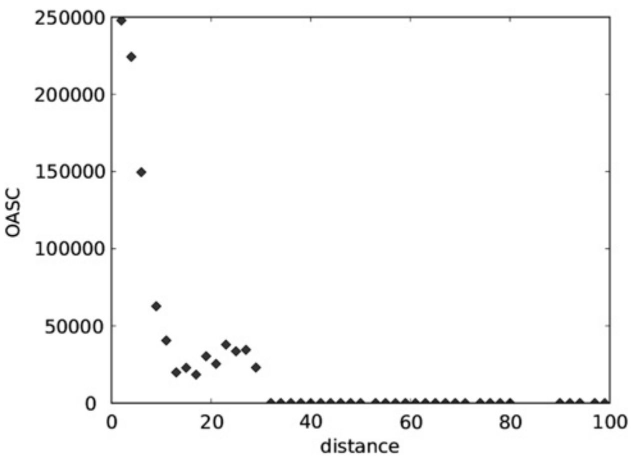


Fig. 11 *OASC values for aerial shot of toy city*

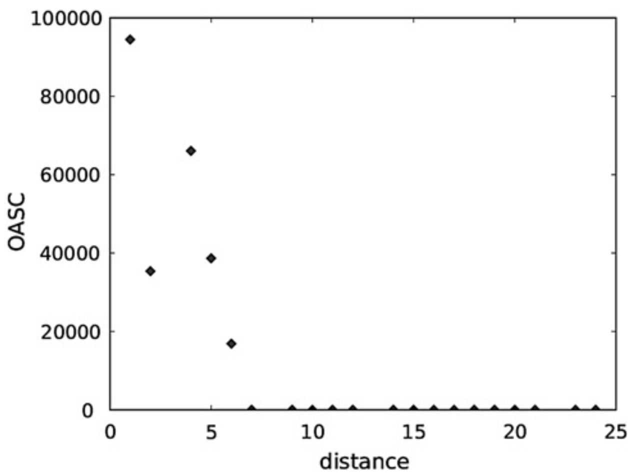


Fig. 12 *OASC values for rocks*

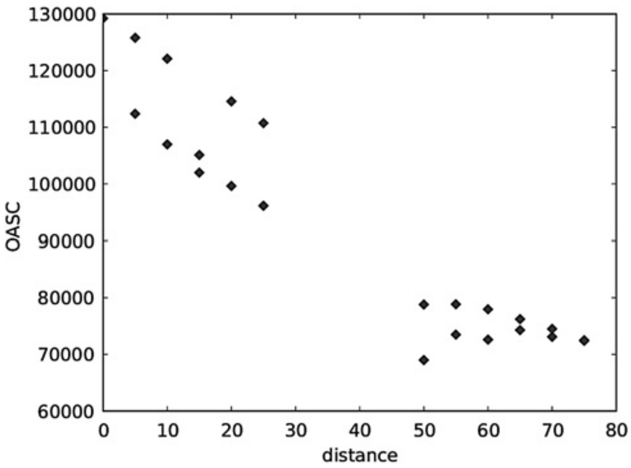


Fig. 13 *OASC values for tiger*

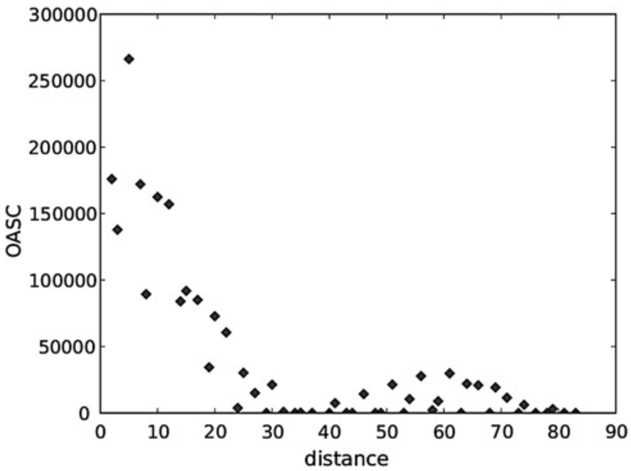


Fig. 14 *OASC values for another toy city*

similar compressions for identical images of Pasteur, and then a picture of Newton placed next to a picture of Pasteur. We obtain the following PNG file sizes

$$K_O(N,\ N) = 77\,\text{kB};\quad K_O(N,\ P) = 148\,\text{kB}$$
$$K_O(P,\ N) = 148\,\text{kB};\quad K_O(P,\ P) = 80\,\text{kB}$$

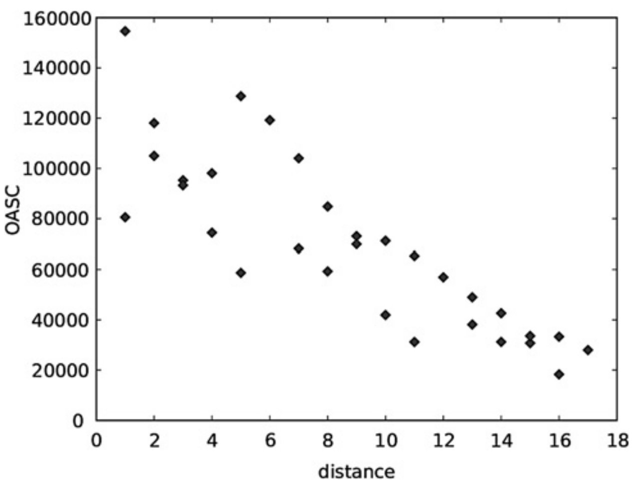


Fig. 15 *OASC values for another toy city (lighter)*

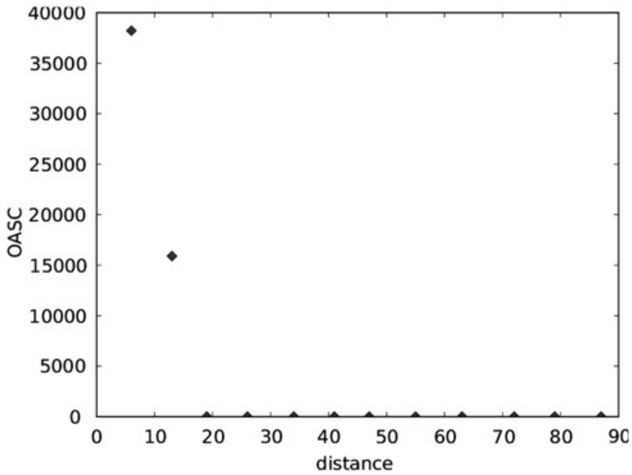


Fig. 16 *OASC values for front shot of city*

There is little redundancy of which to take advantage when the two images are different. The value of $K_O(N, P)$ is therefore larger.

Since [The notation $\overset{O}{=}$ means equality is true up to an additive object-dependent log term, in this case $O(\log K(X, Y))$] [59, 60]

$$K(X,\ C) \overset{O}{=} K(X|C) + K(C) \tag{22}$$

the value of these simple compressed files can be used to estimate the conditional KCS when either the Newton of the Pasteur image is used as context

$$K_O(N|N) = 3\,\text{kB};\quad K_O(N|P) = 72\,\text{kB}$$
$$K_O(P|N) = 74\,\text{kB};\quad K_O(P|P) = 3\,\text{kB}$$

This allows computing of the OASC's using the self-information in (20)

$$\text{OASC}(N,\ N,\ I) = 117\,\text{kB};\quad \text{OASC}(N,\ P,\ I) = 48\,\text{kB}$$
$$\text{OASC}(P,\ N,\ I) = 46\,\text{kB};\quad \text{OASC}(P,\ P,\ I) = 117\,\text{kB}$$

As expected, the OASC of an image of Newton given the same image of Newton as context is very high as is the OASC of Pasteur given Pasteur. Moreover as expected, the cross-cases (Newton given Pasteur and Pasteur given Newton) have a much lower OASC.[Although the relative sizes of the OASC values are most important, cross-term OASC's of 48 and 46 kB are still pretty large. They are a result, in part, of the stochastic model we used for $I(x)$ which will, with probability close to one, give an image of noise. This follows from the asymptotic partition theorem [4, 61]. Moreover, (i) both images have dark backgrounds and (ii) we have not accounted for additive term (from the $\overset{O}{=}$) in (22).]

This simple example illustrates that estimation of ASC can be performed using only the size of compressed files. The applications in data mining are obvious using as context, for example, an ordered bag-of-words [62, 63]. Doing so, although, requires compression that effectively takes into account redundancies that bring the compressed file close to the true KCS complexity. PNG compression and more generally LZW compression works well on shift-invariant [64, 65] (also known as isoplanatic [66, 67], space-invariant or time-invariant) redundancy. PNG comparisons of a shifted versions and Newton next to an unshifted Newton compresses well. Redundancies in shift-variant operations [68–74], such as rotation, scale and transposition, are not captured well by PNG compression. If, for example, the picture of Newton were placed side-by-side with a 90° rotation of the same image, the available redundancy is not taken advantage of by PNG compression. To broadly apply the method of evaluating files using this technique, compression programmes that take advantages of shift-variant redundancy should be used.

3 Conclusion

We have proposed ASC as a methodology to measure the meaning in images as a function of context.

We have estimated the probability of various images by using the number of bits required for the PNG encoding. This allows us to approximate the ASC of the various images. We have shown hundreds of thousands of bits of ASC in various circumstances. Given the bound established on producing high levels of ASC, we conclude that the images containing meaningful information are not simply noise. Additionally, the simplicity of an image such as the solid square also does not exhibit ASC. Thus, we have demonstrated the theoretical applicability of ASC to the problem of distinguishing information from noise and have outlined a methodology where sizes of compressed files can be used to estimate the meaningful information content of images.

4 References

1 Mirowski, P.: 'Machine dreams: economics becomes a cyborg science' (Cambridge University Press, New York, NY, 2002)
2 Marks II, R.J.: 'Information theory & biology: introductory comments', in Marks II, R.J., Behe, M.J., Dembski, W.A., Gordon, B.L., Sanford, J.C. (Eds.): 'Biological information – new perspectives' (World Scientific, Singapore, 2013), pp. 1–10
3 Tribus, M., McIrvine, E.C.: 'Energy and information', *Sci. Am.*, 1971, **225**, (3), pp. 179–188
4 Cover, T.M., Thomas, J.A.: 'Elements of information theory' (Wiley-Interscience, Hoboken, NJ, 2006, 2nd edn.)
5 Bekenstein, J.D.: 'Black holes and entropy', *Phys. Rev. D*, 1973, **7**, (8), p. 2333
6 Hammer, D., Romashchenko, A., Shen, A., Vereshchagin, N.: 'Inequalities for Shannon entropy and Kolmogorov complexity', *J. Comput. Syst. Sci.*, 2000, **60**, (2), pp. 442–464
7 Shannon, C.E., Weaver, W., Wiener, N.: 'The mathematical theory of communication', *Phys. Today*, 1950, **3**, (9), p. 31
8 Grünwald, P.D., Vitányi, P.: 'Kolmogorov complexity and information theory', *J. Logic Lang. Inf.*, 2003, **12**, (4), pp. 497–529
9 Kolmogorov, A.N.: 'Logical basis for information theory and probability theory', *IEEE Trans. Inf. Theory*, 1968, **14**, (5), pp. 662–664
10 Kolmogorov, A.N.: 'Three approaches to the quantitative definition of information', *Problm. Inform. Transm.*, 1965, **1**, (1), pp. 1–7
11 Chaitin, G.J.: 'On the length of programs for computing finite binary sequences', *J. ACM (JACM)*, 1966, **13**
12 Chaitin, G.J.: 'A theory of program size formally identical to information theory', *J. ACM*, 1975, **22**, (3), pp. 329–340
13 Chaitin, G.J.: 'The unknowable' (Springer, New York, New York, USA, 1999)
14 Chaitin, G.J.: 'Meta math!: the quest for Ω' (Vintage, Visalia, CA, 2006)
15 Solomonoff, R.J.: 'A preliminary report on a general theory of inductive inference'. Technical Report, Zator Co. and Air Force Office of Scientific Research, Cambridge, MA, 1960
16 Gitt, W., Compton, R., Fernandez, J.: 'Biological information – what is it?', in Marks II, R.J., Behe, M.J., Dembski, W.A., Gordon, B.L., Sanford, J.C. (Eds.): 'Biological information – new perspectives' (World Scientific, Singapore, 2013), pp. 11–25
17 Oller, J.W. Jr.: 'Pragmatic information', in Marks II, R.J., Behe, M.J., Dembski, W. A., Gordon, B.L., Sanford, J.C. (Eds.): 'Biological information – new perspectives' (World Scientific, Singapore, 2013), pp. 64–86
18 Szostak, J.W.: 'Functional information: molecular messages', *Nature*, 2003, **423**, (6941), p. 689
19 Durston, K.K., Chiu, D.K.Y., Abel, D.L., Trevors, J.T.: 'Measuring the functional sequence complexity of proteins', *Theor. Biol. Med. Model.*, 2007, **4**, p. 47
20 McIntosh, A.: 'Functional information and entropy in living systems' (WIT Press, UK, 2006)
21 Tononi, G.: 'Phi: a voyage from the brain to the soul' (Random House, New York, NY, 2012)
22 Dembski, W.A.: 'The design inference: eliminating chance through small probabilities' (Cambridge University Press, New York, NY, 1998), vol. 112, no. 447
23 Ewert, W., Dembski, W.A., Marks II, R.J.: 'On the improbability of algorithmic specified complexity'. 2013 IEEE 45th Southeastern Symp. on System Theory: SSST 2013, Waco, TX, 2013
24 Ewert, W., Dembski, W.A., Marks II, R.J.: 'Algorithmic specified complexity', in Bartlett, J., Halsmer, D., Hall, M. (Eds.): 'Engineering and the ultimate: an interdisciplinary investigation of order and design in nature and craft' (Blyth Institute Press, Tulsa, OK, 2014), pp. 131–149
25 Ewert, W., Dembski, W., Marks II, R.J.: 'Algorithmic specified complexity in the game of life', *IEEE Trans. Syst. Man Cybern. Syst.*, 2015, **45**, (1), pp. 584–594
26 Stone, W.C.: 'The success system that never fails' (Prentice-Hall, Upper Saddle River, NJ, 1962)
27 Zhu, Q.-F., Yao, W.: 'Error control and concealment for video communication', *Opt. Eng. New York Marcel Dekker Inc.*, 1999, **64**, pp. 163–204
28 Park, J., Park, D.-C., Marks, R.J., El-Sharkawi, M.A.: 'Recovery of image blocks using the method of alternating projections', *IEEE Trans. Image Process.*, 2005, **14**, (4), pp. 461–474
29 Li, M., Vitányi, P.M.: 'An introduction to Kolmogorov complexity and its applications' (Springer, Berlin, 2008)
30 Cyganek, B.: 'Object detection and recognition in digital images: theory and practice' (Wiley, Hoboken, NJ, 2013)
31 Poor, H.V.: 'An introduction to signal detection and estimation' (Springer, Berlin, 1994, 2nd edn.)
32 Thomas, J.: 'An introduction to statistical communication theory' (John Wiley & Sons, New York, 1969)
33 Miller, J., Thomas, J.: 'Detectors for discrete-time signals in non-Gaussian noise', *IEEE Trans. Inf. Theory*, 1972, **IT-18**, pp. 241–250
34 Marks II, R.J., Wise, G., Haldeman, D., Whited, J.: 'Detection in Laplace noise', *IEEE Trans. Aerosp. Electron. Syst.*, 1978, **AES-14**, pp. 866–872
35 Dadi, M., Marks II, R.J.: 'Detector relative efficiencies in the presence of Laplace noise', *IEEE Trans. Aerosp. Electron. Syst.*, 1987, **AES-23**, pp. 568–582
36 Cheung, K., Atlas, L., Ritcey, J., Green, C., Marks II, R.J.: 'Conventional and composite matched filters with error correction: a comparison', *Appl. Opt.*, 1987, **26**, pp. 4235–4239
37 Marks II, R.J., Atlas, L.: 'Composite matched filtering with error correction', *Opt. Lett.*, 1987, **12**, pp. 135–137
38 Marks II, R.J., Ritcey, J., Atlas, L., Cheung, K.: 'Composite matched filter output partitioning', *Appl. Opt.*, 1987, **26**, pp. 2274–2278
39 Vitányi, P.M.: 'Meaningful information', *IEEE Trans. Inf. Theory*, 2006, **52**, (10), pp. 4617–4626
40 Bennett, C.H., Gács, P., Li, M., Vitányi, P.M., Zurek, W.H.: 'Information distance', *IEEE Trans. Inf. Theory*, 1998, **44**, (4), pp. 1407–1423
41 Nikvand, N., Wang, Z.: 'Generic image similarity based on Kolmogorov complexity'. 2010 17th IEEE Trans. on Image Processing (ICIP), 2010, pp. 309–312
42 Supamahitorn, S.: 'Investigation of a Kolmogorov complexity based similarity metric for content based image retrieval'. Masters thesis, Oklahoma State University, 2004
43 Kramm, M.: 'Image group compression using texture databases', in Rogowitz, B. E., Pappas, T.N. (Eds.): 'Human Vision and Electronic Imaging XIII' *Proc. SPIE*, 2008, **6806**, pp. 680513-1–680513-10
44 Lee, J.-D., Wan, S.-Y., Ma, C.-M., Wu, R.-F.: 'Compressing sets of similar images using hybrid compression model'. Proc. IEEE Int. Conf. on Multimedia and Expo, IEEE, 2002, no. l, pp. 617–620
45 Chaitin, G.: 'Kolmogorov complexity and information theory'. Available at http://www.umcs.maine.edu/č haitin/ontology.pdf, 2014, accessed 20 October 2014
46 Ziv, J., Lempel, A.: 'A universal algorithm for sequential data compression', *IEEE Trans. Inf. Theory*, 1977, **23**, (3), pp. 337–343
47 Ziv, J., Lempel, A.: 'Compression of individual sequences via variable-rate coding', *IEEE Trans. Inf. Theory*, 1978, **24**, (5), pp. 530–536
48 Welch, T.A.: 'A technique for high-performance data compression', *Computer*, 1984, **17**, (6), pp. 8–19
49 SureFile, R.: 'Software powered by PKZIP... BSSF DS 0103 authorized reseller: Technical specifications platforms Microsoft® Windows® 98 second edition me | atNT 4.0 workstation sp6a 2000 professional sp2.'
50 Deutsch, L.P.: 'DEFLATE compressed data format specification version 1.3'. Available at https://www.tools.ietf.org/html/rfc1951, 1996, last accessed 15 January 2015
51 Kohno, T.: 'Analysis of the WinZip encryption method', *IACR Cryptol. ePrint Arch.*, 2004, **2004**, p. 78
52 Boutell, T.: 'PNG (Portable Network Graphics) Specification Version 1.0', 1997
53 Roelofs, G., Koman, R.: 'PNG: the definitive guide' (O'Reilly & Associates, Inc. Sebastopol, CA, 1999)
54 Keys, R.: 'Cubic convolution interpolation for digital image processing', *IEEE Trans. Acoust. Speech Signal Process.*, 1981, **29**, (6), pp. 1153–1160
55 Wang, C.-C.: 'Vision and Autonomous Systems Center's Image Database'
56 Costa Santos, C., Bernardes, J., Vitányi, P.M., Antunes, L.: 'Clustering fetal heart rate tracings by compression'. 19th IEEE Int. Symp. on Computer-Based Medical Systems, 2006. CBMS 2006, 2006, pp. 685–690
57 Keogh, E., Lonardi, S., Ratanamahatana, C.A.: 'Towards parameter-free data mining'. Proc. of the Tenth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2004, pp. 206–215
58 Cilibrasi, R., Vitányi, P.: 'Automatic extraction of meaning from the web'. 2006 IEEE Int. Symp. on Information Theory, 2006, pp. 2309–2313
59 Vereshchagin, N.K., Muchnik, A.A.: 'On joint conditional complexity (entropy)', *Proc. Steklov Inst. Math.*, 2011, **274**, (1), pp. 90–104
60 Zvonkin, A.K., Levin, L.A.: 'The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms', *Russ. Math. Surv.*, 1970, **25**, (6), p. 83
61 Dembski, W.A., Marks II, R.J.: 'Conservation of information in search: measuring the cost of success', *IEEE Trans. Syst. Man Cybern. A, Syst. Hum.*, 2009, **39**, (5), pp. 1051–1061
62 Li, T., Mei, T., Kweon, I.-S., Hua, X.-S.: 'Contextual bag-of-words for visual categorization', *IEEE Trans. Circuits Syst. Video Technol.*, 2011, **21**, (4), pp. 381–392
63 Wallach, H.M.: 'Topic modeling: beyond bag-of-words'. Proc. 23rd Int. Conf. on Machine Learning, 2006, pp. 977–984
64 Marks, R.J., Walkup, J.F., Hagler, M.: 'Sampling theorems for linear shift-variant systems', *IEEE Trans. Circuits Syst.*, 1978, **25**, (4), pp. 228–233

65 Martin, J., Baylis, C., Marks, R., Moldovan, M.: 'Perturbation size and harmonic limitations in affine approximation for time invariant periodicity preservation systems'. Submitted to IEEE Waveform Diversity Conf., 2011
66 Marks II, R.J., Krile, T.F.: 'Holographic representation of space-variant systems: system theory', *Appl. Opt.*, 1976, **15**, (9), pp. 2241–2245
67 Marks II, R.J.: 'Handbook of Fourier analysis & its applications' (Oxford University Press, Oxford, New York, 2009)
68 Marks II, R.J., Walkup, J.F., Hagler, M.O.: 'Volume hologram representation of space-variant system', in Marom, E.E., Friesem, A., Wiener-Aunear, E. (Eds.): 'Applications of holography and optical data processing' (Pergamon Press, Oxford, 1977), pp. 105–113
69 Krile, T., Marks II, R.J., Walkup, J.F., Hagler, M.O.: 'Holographic representations of space – variant systems using phase-coded reference beams', in Sincerbox, G.T. (Ed.): 'SPIE selected papers in holographic storage' (SPIE Optical Engineering Press, Bellingham, WA, 1994)
70 Marks II, R.J., Walkup, J.F., Hagler, M.O., Krile, T.F.: 'Space-variant processing of 1-d signals', *Appl. Opt.*, 1977, **16**, (3), pp. 739–745
71 Krile, T.F., Marks II, R.J., Walkup, J.F., Hagler, M.O.: 'Holographic representations of space-variant systems using phase-coded reference beams', *Appl. Opt.*, 1977, **16**, (12), pp. 3131–3135
72 Marks II, R.J.: 'Two-dimensional coherent space-variant processing using temporal holography: processor theory', *Appl. Opt.*, 1979, **18**, (21), pp. 3670–3674
73 Marks II, R.J., Walkup, J.F., Hagler, M.O.: 'Sampling theorems for shift-variant systems'. Proc. of the 1977 Midwest Symp. on Circuits and Systems, Texas Tech University, Lubbock, August 1977
74 Krile, T., Marks II, R., Walkup, J., Hagler, M.: 'Space-variant holographic optical systems using phase-coded reference beams'. 21st Annual Technical Symp., 1977, pp. 6–10