Rensselaer NSF REU midterm report: Quantifying visual information

Alex Striff
Department of Physics, Reed College, Portland OR 97202, USA

Vincent Meunier
Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy NY 12180, USA

(Dated: July 10, 2020)

We attribute a notion of lost information (entropy) to digital images based on considering pixel value fluctuations that are considered imperceptible or irrelevant to understanding the content of an image. This information is precisely defined in the framework of information theory, and is influenced by an analogous situation in statistical physics. Using this analogy enables us to compute the entropy using the Wang-Landau algorithm for the density of states developed for use in statistical physics. Given the results, we discuss potential extensions of the model that better resemble the known physical and biological processes that take place in the visual system.

I. INTRODUCTION

The human visual system is crucial for survival. While this system gathers useful information from our environment, it remains imperfect. A precise understanding of the information lost in sensation may guide future understanding of the information we do gather, as well as inform the design of image compression and reconstruction algorithms for human consumption.

As a simple model for this lossy process, we consider slightly varying the pixel values of a grayscale image to produce different images that are not percieved to be different, or that are not different enough to change the qualitative impression of the image on an observer. The freedom to choose different modified images represents information that is lost. What follows is a description of how we can quantify this information, and then how we can calculate it for particular images.

II. THEORY

A. Information Theory

A mathematical notion of information is provided by the work of Shannon [1]. In the view of classical information theory, *information* is a property of an event, in the sense of a random process. To this end, we consider a random variable X with support $\mathcal{X}$ and probabilities $p(x)$ for $x \in \mathcal{X}$. As regards communication, the information $I(x)$ required to describe the event $X = x$ should satisfy intuitive axioms:

- If $p(x) = 1$, the event is certain to happen and no information is required to describe its occurence: $I(x) = 0$.
- If $p(x) < p(x')$, then x is less likely to happen, and ought to require more information to describe: $I(x) > I(x') \geqslant 0$. As an analogy, compare the phrases “nightly” and “once in a full moon:” The less probable event has a longer description.
- For independent events x and y, it makes no difference to describe the combined event (x, y) instead of each event individually: $I(x, y) = I(x) + I(y)$.

Given these axioms, the only solution is the *self-information* (Theorem A.1)

$$I(x) = -\log p(x), \tag{1}$$

where the base of the logarithm determines the units of information: base two (lg) gives *bits* and base e (ln) gives *nats*. The information of the entire random variable may then be defined as the average of (1) over all events, which is known as the *Shannon entropy*

$$H = -\sum_{x\in\mathcal{X}} p(x) \log p(x). \tag{2}$$

The Shannon entropy may also be derived from intuitive axioms similar to those for the self information [1, 2].

B. The maximum entropy principle (MaxEnt)

A physicist familiar with statistical mechanics might wonder why Shannon’s entropy (2) has the same mathematical form as the thermodynamic state variable for temperature

$$S = -k_B \sum_{x\in\mathcal{X}} p(x) \ln p(x),$$

which we may call the *Gibbs entropy*. This connection between information theory and statistical physics was developed by Jaynes, culminating in the maximum entropy principle (MaxEnt) [2]. We would like to make predictions about systems given some macroscopic quantities that we observe. To do so, we must assign probabilities to microstates, which we ought to do in an unbiased way, subject to the constraints that average macroscopic quantities take their observed values. Jaynes argues that this unbiased assignment corresponds to maximizing the entropy, and describes how this subjective assignment can be expected to make physical

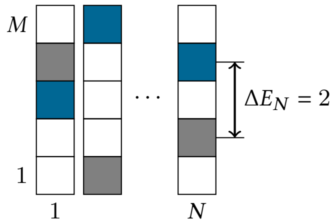

FIG. 1. The difference between base image values and modified image values

predictions, while an objective assignment of probabilities is required to understand the microscopic mechanisms behind these predictions. In particular, maximizing the entropy with constrained average energy produces the canonical distribution [2]

$$p(x) = \frac{1}{Z} e^{-\beta E(x)},$$

where $\beta = 1/k_B T$ and the *partition function* is

$$Z = \sum_{x \in \mathfrak{X}} e^{-\beta E(x)},$$

with the variates x being different states of a system.

C. Image fluctuations

Given a base image x with N pixels which take integer gray values $1 \leqslant x_i \leqslant M$, we define the *energy* of a different image y with gray values $1 \leqslant y_i \leqslant M$ as

$$E_x(y) = \sum_{i=1}^{N} |x_i - y_i|,$$

as depicted in Fig. 1.

We would like to consider all possible modified images, but focus on images with a typical value for the energy which indicates the size of fluctuations we are considering. We do this by assigning a probability distribution to the images with constrained average energy. Given the results of Sec. II B, we choose the MaxEnt distribution, which we may consider as a canonical ensemble. By thinking of our images as a physical system, we may apply tools from statistical mechanics. We would like to know the entropy of the MaxEnt distribution, which we will compute with the partition function as

$$S = \frac{E - F}{T},$$

where E is the average energy we set, T is the temperature, and

$$F = -k_B T \ln Z$$

is the Helmholtz free energy. In turn, we obtain the partition function

$$Z = \sum_{E \in E(\mathfrak{X})} g(E) e^{-\beta E}$$

from the number of states $g(E)$ with energy E (the *density of states*). For the case where the base image is all black ($x_i = 0$) or all white ($x_i = M$), we may explicitly count that the density of states is (Theorem A.2)

$$g(E) = \sum_k (-1)^k \binom{N}{k} \binom{N + E - Mk - 1}{E - Mk}.$$

However, the situation for general grayscale images becomes more complicated. For this reason and the ability to analyze more complex systems, we determine the density of states numerically using the Wang-Landau algorithm [3].

III. METHODS

The Wang-Landau algorithm was implemented to determine the density of states for grayscale image fluctuations.

Appendix A: Theorems

Theorem A.1. *The only twice continuously differentiable function $I(x)$ that satisfies the axioms in Sec. II A is the self-information $I(x) = -\log p(x)$.*

Proof. Consider independent events x and y with probabilities p and p'. The axioms only concern the probabilities of the events, so we may express the information as $I(x) = \tilde{I}(p(x))$. Then as proposed,

$$I(x, y) = \tilde{I}(pp') = \tilde{I}(p) + \tilde{I}(p')$$

by independence. Taking the partial derivative with respect to p gives

$$p' \tilde{I}'(pp') = \tilde{I}'(p),$$

and then taking the partial derivative with respect to p' gives

$$\tilde{I}'(pp') + pp' \tilde{I}''(pp') = 0.$$

We may then define $q = pp'$ to obtain the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}q} \left(q \tilde{I}'(q) \right) = 0,$$

which has solution

$$I(q) = k \log q$$

for real k. The condition that $I(q) \geq 0$ requires $k > 0$, which is equivalent to a choice of base for the logarithm. $\square$

Theorem A.2. *The number of tuples $(x_1, \ldots, x_N)$ with $0 \leqslant x_i \leqslant M - 1$ and $\sum_i x_i = E$ is*

$$g(E) = \sum_k (-1)^k \binom{N}{k} \binom{N + E - Mk - 1}{E - Mk}.$$

Proof. We represent the sum E as the exponent of a integer polynomial in z in the following way. For the tuple (x_1, x_2), we represent x_1 as z^{x_1} and x_2 as z^{x_2}. Together, we have $z^{x_1} z^{x_2}$, which gives the monomial $z^{x_1+x_2} = z^E$ for this tuple. We may then find $g(E)$ as the coefficient of z^E in

$$\left(1 + \cdots + z^{M-1}\right)^N.$$

Expanding using the binomial theorem gives

$$\begin{aligned}\left(\frac{1-z^M}{1-z}\right)^N &= \sum_{k=0}^{N} (-1)^k \binom{N}{k} z^{Mk} \sum_{j=0}^{\infty} (-1)^j \binom{-N}{j} z^j \\ &= \sum_{k=0}^{N} \sum_{j=0}^{\infty} (-1)^k \binom{N}{k} \binom{N+j-1}{j} z^{Mk+j}.\end{aligned}$$

The value of j for z to have exponent E is $j = E - Mk$, so the coefficient of z^E in the polynomial is

$$g(E) = \sum_{k} (-1)^k \binom{N}{k} \binom{N+E-Mk-1}{E-Mk},$$

where the limits of summation are set by the binomial coefficients. □

[1] C. E. Shannon, A mathematical theory of communication, The Bell system technical journal **27**, 379 (1948).

[2] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[3] F. Wang and D. P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states, Phys. Rev. Lett. **86**, 2050 (2001).

[4] Pantin, E. and Starck, J.-L., Deconvolution of astronomical images using the multiscale maximum entropy method, Astron. Astrophys. Suppl. Ser. **118**, 575 (1996).