AN INFORMATION THEORIC IMAGE VISUALIZATION SYSTEM

John S. Powers

NOAA/NESDIS/Interactive Processing Branch
Washington, D.C. USA

ABSTRACT:

The mathematics of information theory may be used to visualize
the metrics of information embodied by digital multispectral
imagery. These visualizations are analogous to the processing
which occurrs at the retinal level of biological visual systems.
The utility of the visualizations lies in their ability to reveal
subtle nuances of phenomena in satellite remote sensor imagery,
and in their shared mathematical foundations with theoretical
studies of systems, complexity and diversity.

KEY WORDS: entropy, image processing, information theory

Introduction

The traditional model of digital imagery
is that of a record of spatially unique,
instantaneous physical measurements within
selected intervals of the electromagnetic
spectrum. From the perspective of
conveyed information at a point, all
points on the image are regarded as
equivalent. Variations on this model
include measurements of deviations from a
datum, such as terrain elevation or
atmospheric pressure.

At a different level of abstraction, an
image may be regarded as an integrated set
of pure information, mathematically
defined. In the integrated model, each
instantaneous value on an image is a
function of all other instantaneous values
over the entire image field. The
abstraction of an absolute physical
measurement at a point is superseded by
the abstraction of information conveyed at
a point relative to an organized system.
This is the model which best describes the
retinal stage of biological imaging
systems.

A robust mathematical definition of
information was first articulated by
Claude Shannon, extending and adapting the
mathematics of classical thermodynamics
(Shannon, 1949). The theory was
originally conceived to solve the problem
of optimizing electronic communications in
the presence of noise over limited carrier
bandwidths, but has since been accepted as
a rigorous mathematical foundation for
broadly applicable metrics of system
organization and diversity.

Organization and diversity are fundamental
characteristics of natural systems,
possessing strong associations with system
behavior, resilance and adaptability to
change. Information theory has fascinated
researchers with its potential as a
paradigm integrating within a single
conceptual and mathematical framework the
diverse fields and scales of inquiry into
systems. The theory has also been
criticized as being frequently
overextended or misapplied to areas where
the mathematical justifications for its
use are tenuous, and the interpretations
of its results questionable.   These
criticisms are well founded, and rigorous
mathematical prerequisites for the
application of the theory must be met.

Discrete Markoff Processes and Ergodic Sources

A fundamental concept of information
theory is that of a discrete Markoff
process. The general case of a discrete
Markoff process is that of a system with a
finite number of discrete states S1,
S2...Sn. For each state of the system
there must exist a set of transition
probabilities Pi(j) which describe the
probability that a system in state Si will
transition to state Sj. A discrete
Markoff process becomes an information
source by producing a measurement at each
transition from one state to another.
Among possible Markoff processes,
information theory concerns a distinct
class known as "ergodic" processes. An
ergodic Markoff process is a process which
generates sequences of values which share
the same statistical properties. As the
length of a sequence from an ergodic
Markoff process increases, the sequence
statistics will cluster around limits
characteristic of that ergodic Markoff
process. Such clustering will exist even
though every sequence may be unique and it
may be impossible to predict the next
state of any sequence from preceding
states. Shannon equates the ergodic
property with "statistical homogeneity."
All human languages are examples of
discrete ergodic Markoff processes. The
phenomenon of "texture" may likewise be
considered an expression of discrete
ergodic Markoff processes in the spatial
domain. Finally, the superposition of
ergodic Markoff processes is itself an
ergodic Markoff process (Moles, 1967), as
is the case with music.

It is a necessary assertion that a
multispectral imaging scanner viewing the
earth meets the definition of a discrete
Markoff process, and that a digital
multispectral image may be regarded as
record of a discrete Markoff process.
This assertion follows from the
observation that transitions of
instantaneous physical measurements from
one state to another in space may be
described by a finite set of transition

probabilities. Likewise, transitions of instantaneous physical measurements from one state to another across spectral channels may be described by a finite set of transition probabilities. This concept is fundamental to the application of information theory to multispectral imagery.

If we model a digital, multispectral image as a record of a discrete ergodic Markoff process, no other mathematical, geometric or qualitative assumptions of any kind are necessary and operations for visualizing the metrics of information theory are exhaustively defined. Spectral, spatial or other statistical components such as illumination intensity or angle which are integrated over the image field are treated identically.

Uncertainty, Entropy and Information

The occurrence of a quantization value or a sequence of quantization values in an image is an event with an associated probability. The probability of an event, and the uncertainty associated with that probability, are separate and distinct concepts. Information theory is founded upon uncertainty and it's measurement. In information theory, the maximum uncertainty possible for any discrete event is log 1/N, where N is the number of possible discrete event states. For an image with a quantization range of 256, the maximum possible uncertainty measurable for any discrete quantization event is log 1/256. For a sequence of events in the same image, the maximum uncertainty possible for the sequence would be $\sum$ log 1/256 for the length of the sequence.

The name given the measurement of uncertainty associated with a set of events is the "entropy." Since maximum uncertainty is conveyed by the least probable event, the least probable event is said to possess the "maximum entropy." Entropy and "information" are commonly confused, resulting in a famous paradox of information theory that only a perfectly random source possesses the maximum entropy or information content. This paradox arises from information theory's use of terminology arising from it's mathematical roots in classical thermodynamics. The paradox disappears if we regard entropy as a measure of the information required to eliminate the uncertainty of an event rather than a measure of information, per se. In some texts, information is referred to as negative-entropy or "negentropy" to distinguish it from the Boltzmann entropy of thermodynamics. This paper adheres to Shannon's original terminology, with the above caveats in mind. Entropy is a relative abstraction devoid of any connection with absolute measurements or criteria. Though simple to compute, it possesses extraordinary conceptual subtlety.

Computing Visualizations of the Information Content of Digital Multispectral Imagery

The first step in the visualization process is to scan the multispectral image and compute two histograms; a raw occurrence histogram and a conditional occurrence histogram. The raw occurrence histogram is a count of the number of occurrences of each quantization value within each spectral channel. The conditional occurrence histogram is a count of co-occurrences of all quantization values across all spectral channels.

The second step in the visualization process is the conversion of the raw and conditional histograms into tables of simple and conditional probabilities which represent the simple and conditional uncertainties associated with image quantization events. It is crucial to make the distinction between an absolute measure of probability and a measure of uncertainty. The maximum uncertainty possible for any quantization event in an image with 256 possible quantization values is 1/256. Therefore, quantization probabilities less than 1/256 must be converted into measurements of uncertainty relative to this probability. For example, probabilities of 1/128 and 1/512 reflect identical uncertainties in an image with a quantization range of 256, i.e., each represents the same amount of information relative to the maximum possible uncertainty. The uncertainty associated with any quantization value is a direct function of the quantization range.

The third step in the visualization process is to re-scan the image and compute visualizations of the metrics of information theory across all spectral channels at each instantaneous resolution element. If spatial entropy is also to be visualized, the computation is summed over nearest neighbor, multispectral spatial samples. Mathematically, visualizations of the entropy of any number of registered spectral channels may be computed but memory and display constraints create a practical limit of three or four channels on small computers.

Metrics of Information: Entropy, Redundancy and Equivocation

Computation of the simple entropy across spectral channels is defined by Theorem 5 (Shannon, 1949b). The simple entropy assumes that probabilities representing uncertainties are independent.

THEOREM 5: Let p(Bi) be the probability of a sequence Bi of symbols from the source. Let

$$Gn = - 1/N \sum_i p(Bi) \log p(Bi)$$

where the sum is over all sequences Bi containing N symbols. Then Gn is a monotonic decreasing function of N and

$$\text{Lim} \quad Gn = H.$$
$$N\text{->}\infty$$

Once the entropy has been computed, the redundancy is defined as one minus the ratio of the measured entropy to its maximum (Shannon, 1949c),

$$\text{Redundancy} = (1 - (Gn / MAX)).$$

Computation of the conditional entropy or "equivocation" across spectral channels is defined by Theorem 6 (Shannon, 1949d). The conditional entropy assumes that probabilities representing uncertainties are conditional upon co-occurring events.

THEOREM 6: Let p(Bi,Sj) be the probability of sequence Bi followed by symbol Sj and pBi(Sj) = p(Bi, Sj) / p(Bi) be the conditional probability of Sj after Bi. Let

$$Fn = - \sum_{i,j} p(Bi,Sj) \log pBi(Sj)$$

where the sum is over all blocks Bi of N-1 symbols and over all symbols Sj. Then Fn is a monotonic decreasing function of N,

$$Fn = NGn - (N - 1) G(n-1),$$
$$Gn = 1/N \sum_{1}^{N} Fn,$$
$$Fn <= Gn,$$

and
$$\text{Lim} \quad Fn = H.$$
$$N\text{->}\infty$$

What Are We Seeing?

Perhaps the most remarkable aspect of visualizations of the metrics of image information is the extent to which the visualizations are easily recognizable images of the natural world. This is in spite of the fact that the metrics of information are mathematical abstractions calculated from tables of uncertainties. Concepts such as "information," "redundant" or "equivocal" do not effectively describe what we are seeing since these concepts imply complex unconscious value judgements which have little meaning relative to the mathematical abstractions visualized. Paradox in the use of familiar language and concepts is inevitable. Probably the best approach on casual inspection is to just view the visualizations as "pictures" which reveal the world in a different light.

In visualizations of image entropy, bright areas represent the "mathematically most interesting" parts of an image, while dark areas represent the "mathematically least interesting" parts. Another interpretation is that bright areas delineate regions conveying the most information about the image field, and dark areas convey the least.

In visualizations of image redundancy, more subtlety of interpretation is involved. Redundancy provides a measure of the amount of information which is pre-determined by context. For example, the redundancy of written English is approximately fifty percent, meaning that almost half of what we write is pre-determined by the structure of the language. In a redundancy visualization, bright areas represent regions which convey the least information relative to the image field. In satellite imagery, the redundancy is dominated by extremely low probability events and is ideal for locating point events which may be spectrally subtle yet statistically prominent, such as fires or hot spots. The redundancy is a logarithmic rather than a linear inverse of the entropy.

The equivocation (or conditional entropy) requires the most subtlety in interpretation. In information theory, equivocation is a measure of information lost during transmission as a function of channel capacity and noise. It represents the ambiguity or residual uncertainty associated with our measurement. Shannon calls it "...the uncertainty when we have received a signal of what was actually sent." In an image visualization, the equivocation seems to represent the coherence of statistical structures. Isolated clouds, for example, generally have very low to zero equivocation values since their multi-spectral statistics are generally very specific (unequivocal!) relative to those of other image components. The equivocation can probably best be thought of as an inverse measure of "clustering" relative to the statistical background. The equivocation occasionally reveals structures absent in both the entropy and the redundancy. Conversely, structures visible in the entropy or redundancy may be absent in the equivocation. Noise in the imaging process is a definite component of the equivocation.

The Entropy Function and Biological Visual Systems

An interesting analog to visualizing the metrics of information in digital multispectral imagery is the processing of the inverted image projected upon the retina of biological visual systems. It must be emphasized that information theory is not a model of retinal processing. Visualizations of the metrics of information theory do share enough characteristics with known retinal operations at the cellular level to make a discussion of retinal processing relevant to their visual interpretation.

In the human visual system, information which reaches the visual cortex in the brain has already been subjected to extensive information processing by the time it leaves the eye. This processing occurrs in the retina, a thin membrane lining the back of the eyeball. The retina is an extension of the neural architecture of the brain and consists of two types of photoreceptors (rods and cones) and several layers of specialized neurons. This layering of distinct neural

structures is a characteristic repeated in the visual cortex (Hubel and Wiesel, 1979). It is believed that each layer of specialized neurons performs specific spatial information processing functions, and that it's output propogates in concert with the functioning of adjacent layers. Visual information processing is an example of what may be called a "cascade" or "pipeline" process, with the retinal transform representing the earliest stages.

Due to the complex spatial interaction among underlying neurons, every photoreceptor in the retina forms the center of a receptive field where light striking the photoreceptor tends to increase the activity of its associated neurons. The activity of these neurons tends to dampen the activity of neighboring neurons associated with other photoreceptors. The effect is mutual, resulting in a phenomenon known as "lateral inhibition." The degree of such spatial interaction between photoreceptors varies inversely with distance, but there is no limit at which it may be regarded as zero. It has been hypothesized that the retina compresses the bandwidth of visual information by suppressing the constant attributes of the visual field while enhancing deviations from the background defined by the whole scene (Mahowald and Mead, 1991).

Visualizations of the metrics of information theory share several interesting similarities with retinal functioning. In both, each instantaneous field of view is a function of the entire visual field as well as stimulation at a point. Both functions replace the bandwidth occupied by the amplitude of stimuli with representations of the relevance of stimuli to a statistical background defined by the entire visual field. Both tend to radically enhance any variation relative to the statistical background. Finally, neither depicts the visual space in an absolute metric. Output remains constant despite variations integrated over the field of view. Conversely, changes in the reference field of view, or of specific structures within it, will result in changes in the representation of everything in the visual field, even things which have otherwise remained constant. This explains how the sudden appearance of subtle events in an otherwise monotonous visual experience can startle us. The retinal function sacrifices the physical objectivity of the visual world by emphasizing those aspects which reveal the most information about the environment. This is the primary strength which the information theoric visualization attempts to exploit.

The Utility of Visualizing the Metrics of Information

Frequently, visualizations reveal coherent structures which are absent or barely perceptible in untransformed imagery. These structures are aspects of the natural world which are being revealed,

however nothing more can be inferred about the structures beyond their relative mathematical "interest" and their spatial structure. Since the mathematical "interest" of an area is a function of the entire image, any change in the spatial context of the image, or any change in the image contents (such as clouds) will potentially change the brightness signature of the structures revealed within it. This extreme relativity can be unnerving for those seeking repeatable, quantitative measurements. Since the structures are real phenomena, their spatial form and location remains constant.

An unfortunate liability with the approach as it now stands is frequent encounters with "granularity," "waves," and "streaking" which arise as the result of noise in the scanner-based imaging systems which contaminates image statistics. Even visually "clean" imagery can contain significant levels of noise which becomes visualized along with other subtle components of the image information field. Noisy imagery can be immediately recognized due to the above effects and inferences derived from the imagery subjected to deserved scrutiny. The potential of this technique with relatively noise-free imagery is an exciting prospect.

Conclusion

A system has been described which enables digital, multispectral imagery to be viewed via the metrics of information theory. Such images possess utility by revealing subtle environmental structures in satellite imagery, and their ability to reveal distortions of the image acquisition process. The images possess a theoretical richness extending from the realm of visual psychophysics to that of the study of natural systems from space. Often useless, occasionally startling and always intriguing, the images reveal the world from the perspective of a seamless mathematical model which begins at the imaging system in space and ends in the brain. They allow us to approach an understanding of the visual and natural world on terms other than a static, cartesian grid of calibrated measurements. These techniques hold exciting research potential into natural systems.

References

Gnedenka, B. and Khinchin, A. 1962. An Elementary Introduction to the Theory of Probability, Dover, New York, 130 pages.

Hubel, D.H. and Wiesel, T.N. 1979. "Brain Mechanisms of Vision", Scientific American, 252(8):150-163.

Khinchin, A. I. 1957. Mathematical Foundations of Information Theory, New York University, New York, 120 pages.

Mahowald, Misha and Mead, Carver. 1991. The Silicon Retina, Scientific American, 264(5):76-82.

Marr, David. 1982. Vision: A
Computational Investigation into the Human
Representation and Processing of Visual
Information, W.H.Freeman and Company, San
Francisco, 397 pages.

Moles, Abraham. 1967. Information Theory
and Esthetic Perception, University of
Illinois Press, Chicago, 221 pages.

Shannon, Claude E. 1949, The Mathematical
Theory of Communication, University of
Illinois Press, Chicago, 125 pages.

Weber, Bruce H. et. al. 1988, Entropy,
Information and Evolution: New
Perspectives on Physical and Biological
Evolution, The MIT Press, Cambridge, 376
pages.