J. C. TRINDER
*University of New South Wales*
*Sydney, Australia*

# Entropy of Aerial Photography

**Information content can be assessed according to the methods adopted for assessing entropy of written text.**

ABSTRACT: *Information theory, as developed by Communications Engineers, defines that the information content, or entropy of a chance event, is a function of the probability that the event will occur. Entropy of a passage of English text can be determined, based on this theory, from the normal probabilities of occurrence of the letters or words in the passage. Here an analogy is drawn between languages, and the different classifications into which photographic information can be subdivided. Entropy of photography is then determined from the probabilities of occurrence of identifiable objects (similar to words in languages) on the photograph. It is proposed that more efficient digital storage may result if digitization is on the basis of the entropy in the photograph.*

## INTRODUCTION

OVER THE PAST DECADE untold volumes have been written on information processing and storage. This has applied mainly to digital information, but recently, the topic of information processing has been extended to aerial photography, which itself is a form of information storage. Felgett & Linfoot (1955), Linfoot (1959, 1961), Altman & Zweig (1963), to name a few authors, have developed mathematical means of assessing a photographic system, as a communication system, based on information theory. Information theory is a product of communications engineering involving, very generally, the problem of information capacity of a communication channel. It has proved a convenient theory also for assessing photographic image quality, because of the similarities between the two systems. All communications channels are perturbed by the general level of noise in the system (e.g., background noise in a telephone). A corresponding effect occurs in photography caused by granularity, introduced by the finite size of the light-sensitive silver halide grains. A discussion of this effect in relation to the general concept of noise, may be found in Linfoot (1965).

A different and less detailed approach than that of Linfoot was adopted by Craig (1962), in which photographic information capacity is analysed in terms of the band width of the spatial frequencies transmitted by the system. Any element in the photographic system (which causes a worsening of the image quality) will mean a loss of the higher spatial frequencies (reduction of the band width) and therefore a reduction in the information-carrying capacity of the system.

Assessment of actual information content of aerial photography to date has involved dividing the photograph into small segments each of which is assumed to have a randomly distributed shade of gray. However, although images are randomly distributed on the photograph, these small segments into which the photograph is divided combine to form identifiable objects on the photograph. The gray shades of each of these segments are correlated with those about them. The assignment of a completely random arrangement of gray shades, therefore, produces an unreal estimate of the information content of the photography.

Information content of photography in this paper will be assessed in a manner similar to that developed by Shannon (Shannon and Weaver, 1949) for the investigation of information content of English text. Such an approach leads to a better understanding of information content of photographs, and may well lead to a more economic digital storage of photographic information.

## Information Theory

### INTRODUCTION

Communications engineers are faced with the problem of transmitting a given message from a source, along a noisy communications channel to a receiver, with neglibible error. The communications channel will always be perturbed by background noise which will have a detrimental effect on the clarity and perhaps even the accuracy of the message. Everyone has experienced background noise in telephones. The received message will be a combination of transmitted message (the signal) and the channel noise. If the noise is large compared with the signal, (e.g., noise strength is approximately equal to signal strength) the result will be an extremely distorted message which the listener may, or may not, understand.

As an aid in solving this problem, a mathematical theory for the assessment of information content of a message has been developed based on the probability of the occurrence of the message. Only the simplest formulas will be introduced for this discussion.

### MEASUREMENT OF INFORMATION—A MATHEMATICAL DEFINITION

Information content of the outcome of a particular event (e.g., the result of the toss of a coin) is a function of the probability of that outcome occurring. For an unbiassed coin, the probability of a head or tail is 0.5, and therefore the result of the toss will contain the same information, whether a head or tail occurs. If however the coin is biassed towards heads, more information will be conveyed if tails occur because a tail is less likely than a head. Likewise, if a person receives a message that "it will be daylight at midday tomorrow", the person will receive very little information, because the event is almost inevitable, and therefore almost of probability one. On the other hand, if the message says that "it will rain tomorrow", the message could certainly contain information. The exact measure of information would depend on the exact probability of it raining tomorrow. Obviously this depends on many climatic factors. Information content is, therefore, a measure of the amount of doubt eliminated by the received message.

A discrete event $S$ has $S_1 \cdots S_n$ possible outcomes, all of probability $p_i$. The information content of the outcome of the event $S$ is (Raisbeck, 1964, 7)

$$H = c \log_2 p_i. \qquad (1)$$

The constant factor $c$ can be chosen such that $H$ is 1 *bit*, for a single event with two equally likely outcomes. The event with 8 equally likely outcomes, therefore, contains $\log_2 8$ or 3 bits of information.

This mathematical expression for information content, or *entropy*, can be extended to evaluate the information content of $n$ independent outcomes $(S_1 \cdots S_n)$ of unequal probabilities $(p_1 \cdots p_n)$ as follows (Raisbeck, 1964, 12):

$$H = \sum_{i=1} -p_i \log_2 p_i \qquad (2)$$

where $\sum p_i = 1$.

The assumptions basic to this definition as outlined by Shannon (Shannon and Weaver, 1949) are:

- $H$ is continuous in $p_i$.
- If all $p_i$ are equal, $p_i = 1/n$, then $H$ should be a monotonic increasing function of $n$.
- If an occurrence is broken down into two successive occurrences, the original $H$ should be the weighted sum of the individual values of $H$ derived from the two successive occurrences.

Messages transmitted by telegraphy can be considered as *events*, as described above. The entropy of telegraphic messages can be analyzed in terms of probability of occurrence of words in the message. Transmitted texts such as "best wishes" contain low entropy because they occur very frequently.

Formulas 1 and 2 are purely mathematical definitions of entropy. They may take account of the receiver's expectation of that message, provided this can be expressed in terms of the probability. These formulas are designed specifically for communications systems, which operate on the principle that frequently occurring phenomena require little information processing, whereas unusual phenomena contain considerable entropy and therefore require more processing. These simple mathematical definitions of information content of a transmitted message are sufficient to determine the entropy firstly of English text in the following sections, and later of aerial photography.

### INFORMATION CONTENT OF WRITTEN TEXT

The basic data transmitted by a telegraphic communications channel via some sort of code (e.g., Morse Code) are words in any number of different languages. For convenience in this paper, only English text will be considered.

Information content, or entropy, of languages can be determined by Formulas 1 and 2, based on the probability of occurrence of letters and words in the language. Each word will have a particular probability of occurrence in a particular language and its entropy can be determined accordingly. The word which appears in two languages having different probabilities of occurrence will have different entropies in each language. Entropy therefore is independent of the receiver's knowledge of the language.

The assessment of entropy of English text, in terms of the common word usage, gives an easy mathematical method of computing information content. However, it does not give an estimate of actual information content, in terms of *meaning* of words, nor does it include the different impacts certain statements may have on people of different backgrounds. Two synonyms one of which is in common use, the other which is not, contain different amounts of information if assessed according to their probabilities of occurrence, e.g., "building" and "edifice". Linguistics is an extremely complex field, and the author will not enter into the discussion of information content in regard to meaning and grammar.

Shannon (Shannon and Weaver, 1949, 43) developed a method of artificially producing English text, based on increasingly refined assumptions on letter and word probabilities. The entropy of texts has been derived using these probabilities by Pierce (1962, Ch. 3) based on a large statistical sample. The different approximations or *orders* of text are as follows:

*i.* Zero-order of English text. Allowing for 26 letters and one space, and assuming equal probability for all letters, a series of "words" can be obtained. For example:

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

Each letter is equally probable, and therefore the information content is given by $\log_2 27 = 4.76$ bits per letter.

*ii.* First-order approximation to English text. The letters are given their own relative probabilities, but each letter is independent of the others:

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAG BRL.

The entropy of this text is 4.03 bits per letter.

*iii.* Second- and (*iv*) third-order approximations can be obtained by adopting digram and trigram (2 letter and 3 letter) probabilities.

*iv.* First-order word approximation is determined by choosing words independently of each other, with their correct relative probabilities:

REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TOEXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

The entropy of this text is 2.14 bits.

*vi.* Second order word approximation includes correct relative and transitional probabilities of words:

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Up to this stage the language still has no meaning, although each higher order of text becomes increasingly similar to English text. As the language becomes more intelligible, the entropy os the text decreases. Even the second-order approximation to English text however, will not give an accurate estimate of the entropy of English text.

Shannon developed a method of evaluating the entropy of actual English text in terms of a person's ability to predict the next letter of a message, given the $n$ ($n$ between 1 and 100) preceding letters. He found that the entropy of English text was approximately 1 bit per symbol as $n$ approaches 100.

It is significant that as the artificially created text becomes more ordered and meaningful and follows prescribed rules of grammar and word usage, the entropy of the text decreases. The orderliness creates redundancy in the text, which means that not all the words and letters in the message are necessary for it to be understood. Consequently, if a message of intelligible English is received containing one or two letters incorrect they can be easily corrected by the receiver from his knowledge of the English language. Weaver (1949, 13) states that the redundancy of English text is about 50 percent. Therefore, half of the letters of English text are controlled by the statistical structure of the language.

Entropy is used to determine the transmission rate or storage capacity required for a particular message, using a given information system. A short discussion on this subject will follow.

## STORAGE OR TRANSMISSION OF ENGLISH TEXT

English text is stored or transmitted for a number of different reasons. One well known case is the transmission of text by telegraph. To commence with, if all letters are assumed to have equal probability, the entropy of the text will be 4.76 bits per letter, and a capacity of 4.76 bits per letter will be required for storage or transmission of this information. This figure assumes a noiseless information processing system.

Assuming the information processing system now includes the entropy of English text, in the form of probabilities of all words, the system will require only 2.14 bits per letter to store or transmit the text. A saving of approximately 50 percent is made in capacity, if the information system includes a knowledge of word probability. Even greater savings may be possible if more sophisticated knowledge of English text is included. However, if the increased knowledge of the text is included, the transmitted or stored message loses some of its inherent redundancy referred to in the previous section. It was pointed out that this redundancy enabled the reader to detect the occasional error which may occur. As the redundancy decreases however, the error detection qualities also decrease. For the most sophisticated method of transmission which would include a complete knowledge of English and an entropy of 1 bit per letter, no redundancy exists, and therefore no error detection is possible.

If the text is stored or transmitted, based on equal probabilities for all letters, the increase in capacity required is offset by the error detection possibilities. Generally the 50 percent increase in capacity required is considered justified (Goldman, 1953, 46). In addition, all languages can be transmitted without difficulty, provided they have the same alphabet. However, for more efficient use of the capacity of the communication channel, letter or word probabilities could be incorporated thereby eliminating some of the redundancy in the text.

## Entropy of Aerial Photography

### INTRODUCTION

In the previous section a purely mathematical approach was adopted for the determination of information content of English text. The influences of meaning and grammar were excluded from the discussion. A similar mathematical approach will be adopted in the following. It should be realized that the entropy computed is a mathematical quantity which will be different from the amount of information that an interpreter may derive from an aerial photograph although the two quantities will not be entirely unrelated.

The simplest method used to determine entropy of a photograph is described by Linfoot (1959, 152). Firstly the photograph can be divided into discrete segments whose size is just larger than the resolution of the image system. Each of these discrete segments on the photograph can have its own shade of gray (for black-and-white photography). The entropy of the photograph can be determined according to the probability of occurrence of each gray shade on the photography. The first choice may be for a completely random pattern of gray shades. Assuming 16 detectable gray shades on the photograph, the entropy of each segment is 4 bits, and the entropy of the whole photograph will then follow. As will be shown heretofore this may amount to some 700 million bits.

Gérardin (1968, 35) suggests that a painting could be divided into small areas, each area being assessed according to a number of variables (e.g., color, texture of grain, density), plus position on the picture, for digitizing the picture. This technique is similar to that described above. Examples of computer generated pictures using elements of black-and-white only, have been published by Julesz (e.g., Julesz, 1964). Pierce (1962, 265) states that the introduction of some degree of orderliness into the pictures improves their appearance, although they may still be completely meaningless.

A more intelligent approach to the assessment of entropy would be to take each identifiable object composed of a number of discrete segments (analogous to words composed of letters in text) e.g., house, road etc., and to calculate its entropy according to its probability of occurring in the region photographed. The size of each object for discussion, is assumed to be such that the image quality of the photograph will not affect its interpretation. The entropy of photography computed from object frequency will be less than the possible maximum computed using randomly distributed gray shades of small segments, because some of the redundancy in the picture has been eliminated. However, this is a more realistic approach for assessing photographic entropy. A user is interested in objects, not gray shades of discrete segments, just the same as a reader is more interested in words than individual letters. This approach will still not give a true estimate of information content. It becomes increasingly difficult to include factors such as links between one object and another, e.g., a road and a house beside it (cf., word connections), or the correct identification of a building or crop (cf., word meaning). Within the limits of the simplified mathematical formulas, however, an estimate of entropy can be obtained.

Information theory was specifically designed for assessing information capacity of

TABLE I

| *Classification* | *Entropy bits/cm²* |
|---|---|
| 1. Cartographic grid | 23.86 |
| 2. Hydrography | 215.04 |
| 3. Relief | 413.30 |
| 4. Populated places | 46.93 |
| 5. Roads | 109.91 |
| 6. Administrative—territorial division | 34.76 |
| Total | 1,025.06 |

communication channels. To this end the entropy of English language has also been investigated. Similarly, the entropy of photographs can be investigated particularly for the purposes of digital storage of photographs, stereo-models and maps. In these instances if the entropy of the photograph, model or map can be investigated, more efficient storage may result.

Digitized information can generally be either (*a*) interpretive and (*b*) metric information. These two topics are discussed in the following sections.

### INTERPRETIVE INFORMATION

Sukhov (1967, 212) has investigated the entropy of a map. To do this, he divided the map into a number of classifications (cf., 'languages' in text) as shown in Table I. Estimates of the entropy of one particular small-scale map under these classifications are also shown. The entropy within each of these classifications was determined, adopting the probability of occurrence of each of the features within the classification. The total entropy of the map was then a sum of entropies of all classifications.

Aerial photography can be divided into similar classifications for interpretive information, e.g.: geology and geomorphology; soils-types; vegetation, timber, crops; cultural improvements.

Within each of these subsections, entropy can be computed in terms of probability of occurrence of various features, for the particular region photographed. For example, in the measurements of terrain form in a geomorphology study in undulating areas, general differences in level of 300 to 400 feet may be expected. Probability for such differences in terrain would be high and therefore they would contain a low measure of entropy. However, a sudden jump of 1,000 feet has a low probability and will convey considerable information to the observer. This is consistent with practical considerations, because such sharp changes in terrain form would have considerable influence on the location of any engineering works in the region.

### METRIC INFORMATION

In order to describe the position of an object relative to another, and relative to an absolute coordinate system, metric detail must be included. For simplified objects (e.g., buildings), this information could be coordinates of the centre of the object. The location of objects on the ground relative to a given coordinate system is generally random, so the digits in a coordinate can be considered as having random probability. Such information could be stored according to existing digital methods.

### APPLICATION OF ENTROPY IN MAP AND PHOTOGRAPH DIGITIZING

As mentioned earlier, one of the possible uses of entropy of photography is in the digitizing of photographs, models and maps. The first instruments designed to digitize photographs have adopted 8 or 16 gray shades for small elements on the photographs (Sharp, 1964; Moore et al, 1964). Such equipment could need $7\times10^8$ bits (or $6.15\times10^5$ bits/cm²) for storage of a 9 inches$\times$20 inches photograph, using 16 shades of gray, and a spot density of $10^6$ per sq. inch. Manipulation of the subsequently formed digitized model of two photographs would be accomplished using a network of points over the model sufficiently dense to describe the form of the terrain. This technique would eliminate some of the redundancy in the photograph. It is noted that the storage required for the photography based on randomly distributed gray shades is considerably larger than the entropy of the map as computed by Sukhov (1967, 212). The two examples are not directly comparable because a map represents only a part of the total information on a photograph. In addition, a map includes place names, contours, etc., which do not appear on the photograph. Nevertheless it is clear that there is substantial redundancy on a photograph, and that digitizing according to identified objects rather than gray shades of small segments would result in greater efficiency.

A model has also been stored by sampling the created model directly at intervals either by a grid or contours (Blaschke, 1968). In this instance considerable sorting of information takes place. The methods used are similar to those used by surveyors on the ground. By doing this a large proportion of the redundancy in the model is eliminated, giving reduced storage requirements. Nevertheless some redundancy still exists, providing a small possibility for error detection. This technique is, in fact, based empirically on entropy of the terrain. For flat terrain with very few changes in grade, if all points in the terrain were sampled, little information would be gained from most of these points, because their heights are predictable. Few sampling points are therefore necessary in flat terrain, which means that only a small storage capacity is needed. If the terrain suddenly changes in grade, as mentioned earlier considerable entropy is implied. A denser network of sampling points in this area is therefore necessary, with a corresponding increase in required storage capacity. Similar considerations should apply if the terrain is sampled by strike and dip (Jackson, 1967).

The entropy of an aerial photograph can be used to digitize interpretive information efficienctly, for example:

* In the interpretation and digitizing of building details on a photograph of a city, entropy can be computed in terms of the probability of size and shape of each building. Such information could possibly be stored in a two-dimensional array. Metric information will be needed to describe the position and orientation (if required) on the photograph. Clearly, there exists an *average* size of buildings which would occur more frequently than the very large or very small buildings. Storage can therefore be done more economically taking these factors into account.
* In tree stands, the size and perhaps shape of the crown can be analyzed and stored accordingly, although investigations into diseases may be possible by analyzing the areas with image tones of diseased trees. The diseased trees may be more significant than the healthy trees and therefore contain more information. Storage may only be required for the diseased trees, whereas all other information can be excluded. Metric information describing the location of such trees must also be included.
* Automated digitizing under classifications described earlier where considerable interpretive skill on the part of an observer is involved, would require highly developed automated equipment not yet available. For example, the classification of tree types or soil types is a question of *meaning* which cannot be assessed by the simplified mathematical rules outlined. However a soil-type or tree-type map produced by the interpreter from the photograph could certainly be digitized. In such maps, the significant features are the borders between tree or soil type classifications and the actual classifications themselves. The borders can be stored by vector information whereas the classification can be described by the requisite number of bits depending on the probability of each classification. Procedures involving vector information are similar to those already accomplished by IBM, as described by Sharp (1968). The number of vectors per inch used to describe a line depends on the curvature of that line. Total storage is allocated according to an *average* expected line curvature, determined after investigation of many maps. This procedure is indeed empirically based on the entropy of lines to be plotted, although it is difficult to say how closely it compares with the true entropy of lines on a map.

The above examples indicate an approach to the digitizing of information in terms of entropy. For analysis of shapes and sizes, pattern recognition equipment is required. Such equipment is already in existence, although only in a simplified form (Gérardin, 1968; Arbib, 1965). For analyses of diseased trees, density or tone equipment required for such a task is in existence. With an inbuilt knowledge of probabilities of occurrence of objects, such equipment could store details as described in a more efficient manner.

## SUMMARY

• The entropy or information content of aerial photography can be assessed according to the methods adopted for assessing entropy of written text. A photograph is made up of many *languages* superimposed on each other. For greater efficiency, digitizing and storage of the photograph could be performed using the entropy of the photography, determined from the probabilities of occurrence of objects in each of these *languages*.

• Automated digitizing however is only possible for simple objects. Details requiring interpretation (or determination of meaning) to date, must be determined by an interpreter before digitizing can take place.

• The approach described in this paper is only a first step in the analysis of entropy of a photograph and subsequent digitizing. A photograph is an efficient storage medium for information although it does contain considerable redundant information. Not all photographic information however, is useful to an individual or institution. Digitizing the photograph using gray shades of small segments is an inefficient method unless all information in the photograph is to be used. More efficient digitizing is certainly possible, particularly where information on one subject only is required, if it is accomplished in terms of entropy within that subject.

REFERENCES

Altman, J. H. and Zweig, H. J. 1963. "Effect of Spread Function on the Storage of Photographic Emulsions," *Photo. Science and Eng.*, Vol. 7, pp. 173–177.

Arbib, M. A. 1965. "*Brains Machines and Mathematics*," McGraw-Hill, New York.

Blaschke, W. 1968. "Problems in Obtaining and Processing of Photogrammetric Data for Digital Terrain Models," *Invited Papef, 11th Congress I.S.P.*, Lausanne.

Craig, R. D. 1962. "Billions of Bits/Minute," *Photogfammet. Eng.*, Vol. 28, pp. 394–406.

Felgett, P. B. and Linfoot, E. H. 1955. "On the Assessment of Optical Images," *Phil. Trans. Roy. Soc.*, 247, pp. 369–407.

Gérardin, L. "*Bionics*," 1968. Translated from French by P. Priban, World University Press, Weidenfeld and Nicolsen, London.

Goldman, S. 1963. "*Information Theory*," Prentice-Hall Inc., New York.

Jackson, V. M. 1967. "Digital Structural Analysis," *Photogram. Eng.*, Vol. 33, pp. 290–296.

Julesz, B. 1964. "Binocular Depth Perception without Familiarity Cues," *Science*, Vol. 145, pp. 356–362.

Linfoot, E. H. 1959. "Information Theory and Photographic Image," *Journal of Phot. Science*, Vol. 7, pp. 148–156.

Linfoot, E. H. 1961. "Equivalent Quantum Efficiency and the Information Content of Photographic Images," *Journal of Phot. Science*, Vol. 9, pp. 188–194.

Linfoot, E. H. 1965. *Fourier Methods in Optical Image Evaluation*, Focal Press, London.

Moore, R. T. et al. 1964. "Digitizing Pictorial Information with a Precision Optical Scanner," *Photogram. Eng.*, Vol. 30, pp. 923–929.

Pierce, J. R. 1962. *Symbols, Signals and Noise*, Hutchison, London.

Raisbeck, G. 1964. *Information Theory*, M.I.T. Press, Massachusetts.

Shannon, C. E. and Weaver, W. 1949. *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.

Sharp, J. V. et al. 1964. "Automatic map Compilation Using Digital Techniques," *Presented Paper, Tenth Congress I.S.P.*, Lisbon.

Sharp, J. V. 1968. "The Influence of Data Processing on Photogrammetric Instruments," *Presented Paper, 11th International Photogrammetry Congress*, Lausanne, Switzerland, 1968.

Sukhov, V. I. 1967. "Information Capacity of a Map. Entropy," *Geodesy and Aerophotography*, (Translated from Russian) No. 4, pp. 212–215.