IS 80% ACCURACY GOOD ENOUGH?

Charles E. Olson, Jr.
Senior Image Analyst
Michigan Tech Research Institute
Ann Arbor, Michigan 48105
ceolson@mtu.edu

ABSTRACT

Before the computer era human interpreters working stereoscopically with black & white aerial photography routinely prepared land cover/use maps with a minimum mapping unit of two hectares and accuracies over 95% for Anderson Level II classes, and 90% for Anderson Level III classes. Today's computer techniques seldom achieve accuracies above 85%, and 80% is becoming an accepted norm. Computers rely primarily, on tone, or brightness. Little success has been achieved in utilizing the other elements of image interpretation used by human interpreters to improve the accuracy of their work. Most computer algorithms assume symmetry, if not normality, in brightness value distributions of terrain targets. Failure of remotely sensed data to meet normality assumptions underlying most classification algorithms is one reason for the seeming cap on land cover mapping accuracy with computers. Cost of computers, operating systems, software, and the technicians required to keep them running, leave few financial resources to do it any other way. This often leads to the mindset that, "If the computer can't do it, it can't be done." While 80% accuracy may be adequate for large regional studies, it may not be adequate to meet the needs of those managing or regulating smaller land parcels.

INTRODUCTION

Accuracy can mean different things to different people, but is almost always a matter of interest. Although geometric accuracy is important, this paper addresses only the accuracy of thematic data, specifically the accuracy of land cover/use information extracted from remotely sensed data. A comprehensive review of this subject for meeting needs of Federal agencies was conducted by an Interagency Steering Committee on Land Use Interpretation and Classification (Anderson, 1971; Anderson, et al., 1971; Anderson, et al., 1976). One conclusion was:

> "The minimum level of interpretation accuracy in the identification of land use and land cover categories from remote sensor data should be at least 85 percent."

This level of accuracy was routinely achieved by human interpreters working stereoscopically with black and white aerial photographs. In one county-wide mapping project using 1:24,000 scale black-and-white aerial photographs and an expanded version of the Anderson Classification System (Anderson, et al., 1971) including more than 50 land cover/use categories (Walton, et al., 1974), accuracies of 96% at Level II and 92% at Level III were achieved. Such results have not been confined to interpretations from traditional aerial photography, for agricultural crop types have been identified with 93% accuracy from a single thermal IR image (Olson, 1967); and combining black-and-white photographs with color-IR photos from NHAP can increase accuracy of land cover determination from 87 to 96% (Olson, 1987). One obvious question is: How do human interpreters accomplish something that today's computer systems don't?

ELEMENTS OF IMAGE INTERPRETATION

Human interpreters integrate several aspects of images to reach their conclusions. These aspects have been defined in several ways (Colwell, 1954; Stone, 1956). The list I summarized 48 years ago was then called the Elements of Photographic Interpretation (Olson, 1960), and is now referred to as the Elements of Image Interpretation. It includes nine EIIs: Shape, Size, Tone, Shadow, Pattern, Texture, Site, Association, and Resolution. The human interpreter, consciously or unconsciously, integrates all of these in arriving at an interpretation; but, most computer algorithms are based almost solely on the tone or apparent brightness of the image.

In addition to integrating the EIIs, interpreters should carefully integrate the many aspects present in their images. Image interpreters, even in our computer age, should be terrain conscious in the broad sense of terrain. They should understand that the images being interpreted portray a terrain that is a complex series of interrelated features academically catalogued as agriculture, botany, engineering, forestry, geology, etc. The importance of these interrelationships was clearly expressed by C. H. Summerson when he said:

> "The subject matter of the photo interpreter is as broad as the earth itself; … upon the earth's surface is a tremendously varied pattern of natural features, upon which man has superimposed an equally varied cultural pattern. … Where there are so many varied objects to be considered, our first inclination is to classify. Classification is indeed necessary, but it must always be remembered that groupings are man-made and not natural. … Conclusions based on empirical identifications of isolated elements in a terrain may often be inaccurate. … The knowledge that no region is a static thing greatly aids the interpreter in his understanding of the terrain." (Summerson, 1954)

Failure to give adequate consideration to all aspects of a terrain is a major cause of misinterpretations. D. J. Belcher put it this way during a discussion of correlations between man's activity and his environment:

> "… the average interpreter does not know many things and he doesn't realize his lack of knowledge. If he places two elements of the environment together and pulls a boner, it isn't because the environment is wrong; instead, it's because there were three other elements that he ignored or didn't know were there." (Belcher, 1957)

Individuals with the breadth of terrain consciousness and integrative abilities identified by Summerson and Belcher are relatively rare. This has led to many attempts to replace the human interpreter with an automated, computer-based system. The difficulties of capturing and reproducing the analytical processes used by human interpreters are not trivial (Hoffman, 1987). Progress over the last two decades has been impressive, but automated interpretation systems seldom achieve classification accuracies above 80% even when dealing with less than twenty terrain classes. Despite this, much of the user community seems satisfied with this lower accuracy. Does this acceptance imply that Anderson et al. (1971 and 1976) were wrong when they specified 85% accuracy as the minimum acceptable level, or is this acceptance based on the belief that 80% accuracy is as good as you can get?

FACTORS AFFECTING ACCURACY

In the early 1970s, National Land-Use legislation was being discussed and it was often assumed that to receive Federal funding, it would be necessary to use the "Anderson Classification System." If one wants to create a National data base, it seems logical that everyone should use the same classification system. It is also reasonable to expect some minimum level of accuracy in the data input to the National data base. When the 85% figure was proposed (Anderson, et al., 1971), then current experience indicated this should be readily achievable. In effect, the 85% figure represented a "lowest common denominator" approach.

The rapid growth of computer-based interpretation systems since the mid-1980s appears to have changed some fundamental assumptions. The 85% figure has been replaced by something lower, essentially 80%. Since consumers are willing to accept and pay for this level of accuracy, it would seem that 80% is good enough.

Concurrent with the improvement in automated interpretation, has been the rapid growth of geographic information systems as the basis for storing and manipulating land cover/use data. As the power of these systems has increased, the cost of software and personnel to keep it all running has also increased, even as computing power costs have decreased. In some cases, software and personnel requirements have taken so much of the budget that there isn't anything left to do it any other way. This can lead to the belief that there "isn't any other way to do it," and as long as the product fits neatly into the GIS data base, it's good enough. Additionally, computer-based classification methods are what students now learn, believe to be the best approach, and are told is required in the workplace. Less and less consideration of manual but more accurate image interpretation methods is being given in training programs and manual image interpretation skills are steadily disappearing from the work force.

It is noteworthy that some jurisdictions still believe manual interpretation methods produce the best results. When Napa County, California needed an accurate county-wide land cover layer, their new layer was "produced by on-screen digitizing over USGS Digital Orthophoto Quarter Quads" by a University of California research team (Thorne et al. 2004). They were able to create a land cover layer that had more vegetation classes than available computer classification-based layers, with finer spatial resolution and improved mapping of wetland areas.

The Normality Assumption

Statistical packages underlying most automated interpretation systems include the assumption that brightness values recorded at the sensor are normally distributed for any given terrain class. This assumption is not true (Olson and Ma, 1989). The distributions are skewed, and when some multiple of the standard deviation is used to establish the decision gates - that range of values in a spectral channel to be accepted as belonging to any terrain class - truncation of the distributions leads to omission errors in that class and possibly commission errors in another class. Tests of a classifier using asymmetrical decision gates with Landsat data have yielded accuracies of 93% without any preprocessing of the data (Olson, 1992). This suggests that as much as 15% of the error in automated classification systems may stem from the failure of the normality assumption. Effective use of the asymmetric decision gate approach requires the kind of broad terrain consciousness referred to by Summerson (1954).

The Time Factor

It often takes human interpreters longer to complete an interpretation product than can be achieved with an automated computer-based system. Except in emergency situations, the speed at which decision processes move forward seldom requires this speed. In effect, satisfying the desire for speedy results comes at the expense of thematic accuracy.

Time may enter the equation in another way. A study of land use change in 79 PLS sections at Phoenix-Tempe, Arizona, is illustrative (Olson, 2006). Data were desired for each 40-acre block, in each of the 79 sections, for four different years: 1949, 1953, 1964, 1999. Those years represented 10-15 year intervals with the specific year dictated by the availability of suitable aerial photographs. For each date, the proportion of each 40-acre block that was in one of five land cover classes - irrigated agriculture, irrigated small urban, irrigated large urban, open water, and non-irrigated, and the proportion of the block that was permeable were desired. At first blush, this looks like a task designed for a GIS solution. To meet the need through the GIS approach, however, it would be necessary to map the entire 79 square miles on each date and show every building, road, street, driveway, sidewalk, swimming pool and patio. Using old fashioned methods the task, including planning, test of the procedures, interpretation of all 1264 40-acre blocks on each date, and accuracy assessment, was completed in ten months with an accuracy above 95% and a total investment of 97.3 man hours, or 1.25 hours per square mile. Had all of the data been available at the beginning of the project, the work could have been completed in four months at the same cost. I have asked several potential vendors to provide an estimate of what they would have charged to complete the same task, but none has been willing to provide even a ball-park estimate, although they did say they could have completed the job more quickly. Would that speed have justified higher cost and lower accuracy?

HOW MUCH ACCURACY IS REALLY NEEDED?

There can be no single answer to the question in this heading. Accuracy needed for a National land-use data base is not the same as that required at the local level. The kind of broad-brush land-use decisions made at National or Regional levels are not the same as those required of zoning boards. One community that has recently enacted a storm-water management act is discovering that 80% accuracy in determining the amount of impervious area on individual land parcels - the basis for determining the amount of the storm-water tax for each parcel - can lead to vociferous complaints from those owners whose parcels have been incorrectly evaluated.

All that is clear is that the present acceptance of 80% accuracy is resulting in poorer data being delivered to decision makers than was routinely delivered, somewhat more slowly, by human interpreters in the five decades since the end of World War II.

LITERATURE CITED

Anderson, J. R., 1971. Land use classification schemes used in selected recent geographic applications of remote sensing. *Photogrammetric Engineering*, Vol. 37(4):379-387.

Anderson, J.R., E.E. Hardy, and J.T. Roach, 1971. A land-use classification system for use with remote-sensor data. U.S. Geological Survey Circular 671, 16 pp.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, 28 pp.

Colwell, R.N., 1954. A systematic analysis of some factors affecting photographic interpretation. *Photogrammetric Engineering*, Vol. 20(3):433-454.

Belcher, D.J., 1957. Remarks made during discussion of a paper by G. R. Heath (see below) and published in *Photogrammetric Engineering*, Vol. 23(1):114.

Heath, G.R., 1957. Correlations between man's activity and his environment which may be analyzed by photo interpretation. *Photogrammetric Engineering*, Vol. 23(1):108-114.

Hoffman, R.R., 1987. The problem of extracting the knowledge of experts. *AI Magazine*, Vol. 8(2):53-67.

Olson, C.E., Jr., 1960. Elements of photographic interpretation common to several sensors. *Photogrammetric Engineering*, Vol. 29(6):968-971.

Olson, C.E., Jr., 1967. Accuracy of land-use interpretation from infrared imagery in the 4.5 to 5.5 micron band. *Annals of the Association of American Geographers*, Vol. 57(2):382-388.

Olson, C.E., Jr., 1987. Further results with NHAP photographs. In, *Color Aerial Photography and Videography in the Plant Sciences and Related Fields*, Proceedings of the Eleventh Biennial Workshop on Color Aerial Photography in the Plant Sciences, Weslaco, Texas, pp. 126-130.

Olson, C. E., Jr. and Z. Ma, 1989. Normality assumptions in supervised classification of remotely sensed terrain data. In, *Quantitative Remote Sensing: An Economic Tool for the Nineties.* IGARSS '89. IEEE #89CH2768-0, pp. 1857-1859.

Olson, C.E., Jr. 1992. Minimizing classification errors arising from skewed distributions of satellite-observed brightness values. In, Proceedings of the ASPRS/ACSM Global Change Symposium, Washington, DC, Aug. 3-8, 1992. Vol. 4:116-121.

Olson, C.E., Jr., 2006. Land-use change inputs to ground water monitoring. In, *Proceedings of the ASPRS Annual Convention*, Reno, NV, May 2006.

Stone, K.H., 1956. Air photo interpretation procedures. *Photogrammetric Engineering*, Vol. 22(1):123-132.

Summerson, C.H., 1954. A philosophy for photo interpreters. *Photogrammetric Engineering*, Vol. 20(3):396-397.

Thorne, J.H. 2004. A vegetation map of Napa County using the Manual of California vegetation classification and its comparison to other digital vegetation maps. *Madrono*. Vol. 51(4):343-363.

Walton, J. M., C. E. Olson, Jr., and E. Limoges. 1974. Mapping land-use from existing air photos. In Proceedings of the 40th Annual Meeting of the American Society of Photogrammetry. St. Louis, Missouri, March 1974, pp. 290-300.