

## **USING GEOREFERENCED LARGE-SCALE AERIAL VIDEOGRAPHY AS A SURROGATE FOR GROUND VALIDATION DATA**

Dana Slaymaker\*

*University of Massachusetts, Department of Natural Resources Conservation, University of Massachusetts, Amherst, Massachusetts, 01003, USA*

### **1. INTRODUCTION**

When mapping forest regions and vegetation from satellite imagery or small-scale photography, it is essential to obtain an adequate sample of geographically distributed, unbiased verification and validation data to both drive the classification and assess the accuracy of the results. Traditionally, this has required on-site visits or “ground truthing” of a randomly selected set of locations distributed across the region to be mapped, an often expensive and time consuming process.

In 1989, the Gap Analysis Program (GAP) of the United States Geological Survey (USGS) began mapping biodiversity across the contiguous United States. The purpose of the program was to identify areas of unique species richness that were not presently covered by the existing network of protected lands controlled by state or federal agencies (Scott and Jennings 1998). This required an up-to-date, detailed vegetation map as a substrate for habitat analysis. For this purpose, the program made the largest single purchase of Landsat 5 satellite images at that time, a complete multi-temporal coverage of the United States (MLRC program). The classifying and mapping of this data involved government agencies and universities in every state and resulted in new methodologies for GIS (Geographic

\* Current affiliation: Winrock International, Ecosystem Services Group, 22 Third Street, Turners Falls, Massachusetts, 01376, USA

Information Systems), GPS (Global Positioning Systems) and aerial imagery that have permanently changed the way these tools are used in resource inventory.

One product of that revolution was an alternative approach to ground truth verification which used the interpretation of very large-scale georeferenced aerial videography. The development of a time-coding device to automatically label video frames with Global Positioning System (GPS) co-ordinates in the 1980s (Graham 1993) made it practical to extend an aerial point sampling technique originally developed for 35 mm photography (Norton-Griffiths 1988) to extensive regional surveys, collecting large numbers of samples inexpensively. The original technique was based on the accepted premise that photographic interpreters can be trained to accurately identify trees and vegetation at a sufficiently large scale to detail individual crown structures (Sayn-Wittgenstein 1978; Drake 1996). By flying grid patterns of coverage across a region, thousands of georeferenced frame samples can be collected very quickly. Photographic interpreters are then trained on a small selected number of those sites within each forest stratum class. The sites are visited and trees in each video frame identified directly on a print of the image. Visual indexes are made of these prints, which identify each tree species or forest type within that specific set of coverage. Using those visual indexes to interpret the rest of the frame samples results in a much richer set of known ground reference points that can drive the classification of a satellite image or verify any interpretation made at a much smaller scale (Slaymaker et al. 1996).

Since the 1980s, aerial videography has seen increased use in applications where its advantages over traditional photography (lower cost and immediate availability of data) outweigh its disadvantages (poorer spatial resolution and difficulty of analysis due to lack of stereo imaging) (Mausel et al. 1992; Meisner 1986). King (1995) provides a comprehensive review of the evolution of video sensors and their applications, many of which focused on: 1. The measurement of transient phenomena such as wildlife populations (Sidle and Ziewits 1990; Strong and Cowardin 1995) and pest infestations (Everitt et al. 1994); 2. Mapping of dynamic land features such as wetland plant communities (Jennings et al. 1992) and coastal landforms (Eleveld et al. 2000); 3. Land cover mapping in remote areas with limited existing aerial photography and poor infrastructure (Marsh et al. 1994; Slaymaker and Hannah 1997).

## 2. METHODS

The basic technique of flying a grid of large-scale images to classify Landsat imagery was originally developed in 1980 by a team of researchers (Dunford et al. 1983) for several USAID projects in Africa. Using 35 mm film, vertical or oblique exposures were taken from 300 to 1000 feet above ground. Each exposure was fired by hand and manually georeferenced by reading co-ordinates off a Global-Nav system at the time of exposure.

Arizona was the first state Gap Project to try to use this modification of Norton-Griffiths's point sampling approach to classify the Landsat coverage. In spite of its lower resolution and poorer colour quality, video was chosen over 35 mm film because a GPS-based SMPTE (Society for Motion Picture and Television Engineers) time code generator had been developed by the Horita Corporation. Devices that wrote GPS co-ordinates directly onto video frames through a caption generator had been used before (Myhe et al. 1991), but these records were only accurate to the last one-second position. The Horita wrote the time code and frame number to every exposure, so that the geographic position of each frame could be calculated. Only one 35 mm camera, the Nikon F4 with a 250 exposure back, could label images to the nearest second provided its internal clock was manually matched to GPS time, and this was insufficient for the scale at which the coverage was to be flown. Cost and in-flight management of data were also major considerations as evident in the Tables 18-1 to 18-3.

*Table 18-1. Cost of digital images for each system by data collection time*

Time and media	Cost (US\$)
Two hours of aerial videotape	\$7.50
Two hours of 35 mm camera slides, 2 exposures every 10 seconds (stereo pairs) scanned to 2,000 by 3,000 pixels by Kodak and stored on CD.	\$2,085

*Table 18-2. Cost of commonly used camera and video equipment in 1990.*

Camera	Cost (in US\$)
Panasonic Super VHS camcorder	\$1,200
One Nikon F4 with 250 exposure back	\$8,000

*Table 18-3. Intervals required to change film or tape*

Camera	Time
Video camera	Every two hours
35 mm camera with 36 exposures.	Every 6 minutes
35 mm camera with a 250 exposure back	Every 50 minutes

Arizona is covered by 16 Landsat images in 5 tracking swaths. Video was flown over the state in a 30-kilometre grid using a Panasonic Super VHS camcorder and 12X zoom lens mounted vertically in an aerial photo aeroplane with a standard Fairchild camera mount. The camera was flown approximately 2,000 feet above ground, manually zooming to telephoto every 9 seconds, so as to provide both a wide-angle view of the terrain and a set of large-scale samples. Several thousand video points were collected over the state, but the total sample averaged about 600 ground reference points per Landsat image. This proved to be insufficient to successfully model the mixtures of vegetation and soil types within the range of slope/aspect variations that affected their spectral reflectance. The principal reason for the low number of sample points per image was the use of a single camera alternating between the wide angle and zoom settings, drastically reducing the number of large-scale images that could be used for sampling. In subsequent surveys, that system was replaced with two Hi8 video camcorders that were less expensive and could be mounted to the window of any Cessna aircraft. This arrangement allowed for much a denser sampling because the telephoto camera provided a continuous large scale swath within the wide-angle view, rather than an occasional zoom (Figure 18-1).

This technique was first used successfully to classify the forests of the North-eastern United States for the Gap projects of Maine, New Hampshire/Vermont and Southern New England (Massachusetts, Connecticut, and Rhode Island). Mapping vegetation in this part of the world represented some unique problems; the landscape is 50 % to 95 % forested with a wide variety of forest types occurring in relatively small stands interspersed in a pattern that does not follow any readily discernible rules. While most vegetation types in the United States have developed along natural limits and can often be modelled with respect to variations in elevation and terrain, the chief determiner of species distribution within the major forest regions of New England is historic human activity. Over 80 % of Massachusetts was agricultural land in the 1800's and has since been replanted or has re-grown to forest. The resulting landscape is an intricate mosaic of different forest communities arranged in a whimsical fashion. In addition, the initial examination of unsupervised classifications of the Landsat coverage available for the area showed a less than direct relationship between their classes of spectral reflectance and the SAF (Society of American Foresters) forest types that were to be classified. Most forest types were composed of mixtures of spectral classes. Few spectral classes represented more than 15 % to 35 % of any one forest type and most forest types contained most spectral classes to some degree. This spectral mixing is the essential problem of any image analyst trying to regroup spectral classes

into forest types. They are fundamentally different kinds of categories, one based on the reflectance of energy and the other on human perceptions of what constitutes a community of plants.

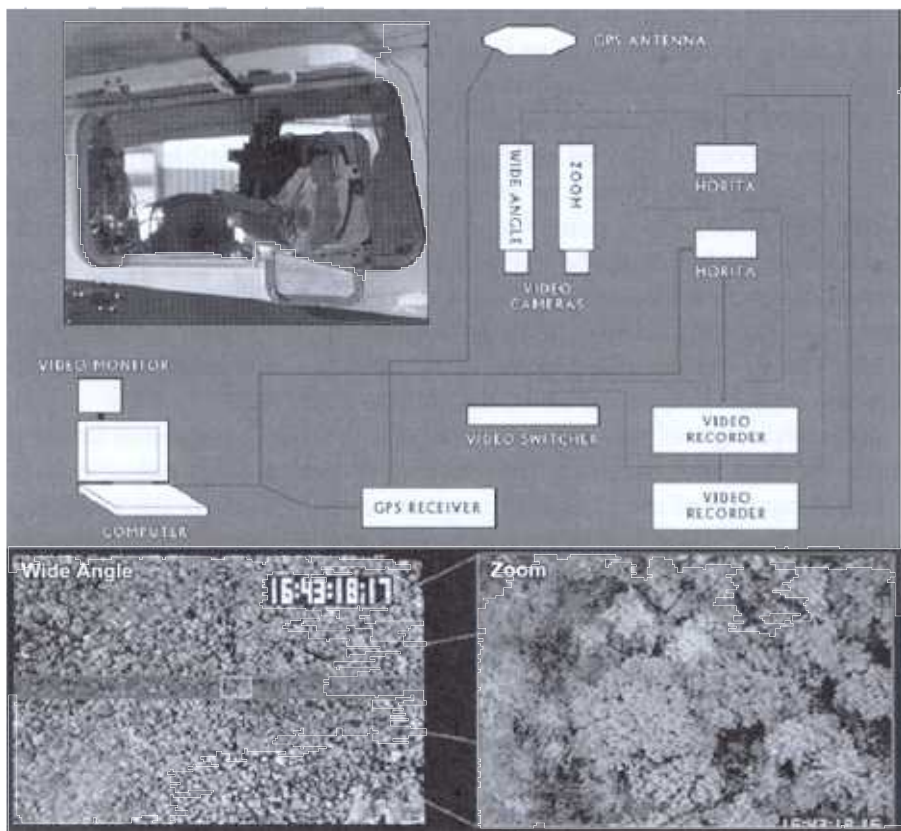


Figure 18-1. Dual camera Hi8 video system mounted in the window of a Cessna 172. A GPS receiver sends data to two Horita SMPTE time code generators, which write the time and frame number to every exposure. Flying at 2,000 ft. above ground, the wide-angle camera covered a 500-meter swath at one meter per pixel resolution. The zoom camera sub-sampled the centre portion of that swath at a resolution of 0.08 meters per pixel.

To deal with this complexity, these projects were much more densely sampled than the Arizona Gap Analysis coverage, flying grids lines 15 kilometres apart and collecting approximately 18,000 points per Landsat image in a stratified sampling to capture all vegetation types in all slope/aspect classes. Once the video was collected and reviewed, a sub-sample of sites that were representative of the different vegetation types and accessible by road was selected and visited by ground crews, who used prints from the video coverage of each site to label its dominant trees and

vegetation. These prints were collected into visual indexes used to train the photographic interpreters, using methods developed by Norton-Griffiths and used successfully in Arizona (Drake 1996) (Colour Plate 19).

Points were initially collected by superimposing the GPS flight log over each Landsat classification in an image processing program and viewing the video monitors beside the computer. Two video recorders were keyed to the same remote control, which kept them in relative sync during viewing. Advances in computer-based video eventually made it possible to display the video on-screen and rotate it to match the orientation of the satellite image (Colour Plate 20). Each point was collected by clicking on a pixel in the Landsat image that could be visually identified as belonging to a specific vegetation type on the corresponding video images. This brought up a point panel automatically labelled with X, Y (latitude, longitude) and a blank Z value. The interpreter filled in the Z value with a numerical code for the selected vegetation type and saved the point to a text file.

The advantage of this approach was that a visual comparison could be made between the wide-angle video images and a false colour-infrared colour composite of each Landsat image to correctly identify the location of individual trees in the corresponding zoom video. Because the GPS position of the aircraft could be anywhere within a nine pixel block of confusion when projected to the ground, having a visual image of the corresponding Landsat image made it possible to estimate the most probable location of a video frame within that context, which improved the accuracy of point identification. Slope/aspect classes from the topography of the region were also projected on screen to increase sampling in the less frequent terrain types. Sampling and labelling a point on screen took only a few seconds once the interpreter was familiar with the forest types; so collecting 18,000 points per Landsat image was not as onerous as it sounds.

The Landsat data used in Southern New England had been "hyperclustered" to 240 classes using a classification algorithm developed by the Los Alamos Laboratory in New Mexico (Kelly and White 1994). This type of clustering represented a considerable improvement in the discrimination of spectral classes, but increased the complexity of regrouping them into information classes. Therefore, the essential problem was how to use the thousands of collected ground reference points for each of these classifications to model a re-labelling procedure. It was necessary that this procedure considered both the spatial context of each pixel in the image and the mixture of spectral signatures associated with each forest type, then determined the most probable identity of that pixel. Richards et al. (1982) presented a methodology for regrouping a non-Gaussian unsupervised classification to information classes by incorporating

neighbourhood values into a probabilistic label relaxation procedure. There have been several variations on this method since (Stumpf and Koltun 1992). Southern New England Gap Analysis followed their basic concept, but implemented it through a set of hierarchical inference rules in GRASS (Geographic Resources Analysis Support System), an open source GIS program that is freely available on the web.

GRASS was used for this re-labelling because of two specific routines in the program: *s.menu* and *i.infer*. *S.menu* imported the list of ground reference points as a site file, then generated a list of values for each location from a stack of GIS layers. The layers included the Landsat classification, all ancillary layers (slope/aspect, ecological region, distance from water, etc.) and a number of neighbourhood matrix functions within a 25-pixel block (mode, max, min, diversity, etc.). In effect this procedure placed each point within the context of its surrounding spectral classes and geographic location. These lists of attributes for each reference point were then organised in contingency tables that identified several potential vegetation types for each spectral class. As stated, each vegetation type was seen as being made up of a mixture of spectral classes. Some spectral classes in that mixture would be significant indicators of a specific vegetation type in a particular set of slope/aspect and ancillary factors, and in association with other spectral classes within the immediate neighbourhood of each pixel. However, those same spectral classes could indicate a different vegetation type in a different set of associations or appear in vegetation clusters where their presence was not significant. Therefore, to label mapping units of information classes from the mixtures of spectral classes in the hyperclustered classification, it was necessary to identify each cluster of pixels that represented a specific vegetation type by its significant spectral class members, then absorb the non-significant pixels in those clusters into the same classifications.

Of the 240 spectral classes in the unsupervised hyperclustered classifications of multi-temporal Landsat images used in the Southern New England Gap Analysis, 180 were identified as relevant to vegetation. The net result of attributing these pixels with neighbourhood and slope values in GRASS was to increase the potential sets of spectral/ancillary signatures from 180 to over several thousand. This is why it was necessary to collect so many thousands of ground reference points from the video. They provided a sufficient population of samples to statistically model relationships between those spectral classes in the different contexts provided by the ancillary data layers. As these potential context classes were identified from the contingency tables, they were written into three successive sets of decision

rules in the GRASS program "i.infer", which were run in a cascading fashion in a shell script as follows:

- Rule # 1 – Queried each pixel in the original classification and all underlying 5-pixel by 5-pixel neighbourhoods / ancillary data layers with a series of yes/no - if/and if questions. Pixels that could be assigned to a probable initial vegetation type were assigned; all others were labelled as zero. This rule set ran only once.
- Rule set # 2 – Queried each pixel in three layers, the original classification of spectral classes: the product of rule set # 1, and a 3-pixel by 3-pixel majority of that product (re-labelling each classified pixel as its most frequent neighbour). These rules only looked at those pixels in the Rule # 1 product whose majority vegetation type was different than the type they were assigned to. If that majority vegetation type was one to which the original spectral class of that pixel could belong to, according to the first rule set, then it was relabelled to that majority. Otherwise, it was labelled zero. The product of this rule set was then merged with the product of the first rules, re-labelling all non-zero results. This second rule set was repeated in a loop until the immediate product (those pixels that were reassigned) reached a minimum threshold of non-zero pixels, indicating that little further change would occur.
- Rule set # 3 – Queried each pixel in three layers: the original classification of spectral classes, the final product of Rule set # 2, and a 3-pixel by 3-pixel majority of that product. These rules only looked at those pixels in the final Rule set # 2 product whose value was still zero. If those pixels had a majority vegetation type and it was a type to which the original spectral class of that pixel could belong to, then it was relabelled to that majority. Otherwise, it was left as zero. The product of this rule set was then merged with the final product of the second rule set, re-labelling all non-zero results. This rule set was also repeated in a loop until the intermediate product of the rule reached a minimum threshold of non-zero pixels.

The final product of Rule set # 3 was then run through a series of calculations ("a" if "x" not zero, 0 otherwise) in the GRASS program "r.mapcalc", in which "a" was the 3 by 3 majority of the product and "x" was a mask for all pixels in the original spectral classification associated with vegetation. This was also repeated until a threshold of no change occurred.

The purpose of this complicated cascade of decision rules was to correctly identify kernels of pixels within the natural clusters of vegetation types. It would then grow them towards each other to create the kind of

artificial boundaries that are not found in nature but are considered necessary for thematic classifications with minimum mapping units. This approach would be impossible without the large number of ground reference points that can be economically acquired with aerial videography.

Since only a small fraction of the video frames are actually used in driving the labelling process, the data source can also be used to verify the results. This was originally done in the Southern New England Gap project by setting aside a random subset of the interpreted points (stratified to cover all vegetation types) before generating the contingency tables to model the rules. Those points were then used in an error matrix between their interpretation and the labelled vegetation type. In later surveys, this procedure was changed to sort all the GPS positions in each flight line of the survey grid by their classification in the final map, then select a stratified random subset. These points in the video coverage would then be examined by a new set of photographic interpreters (other than those who had done the original classification) judging the associated vegetation type as: right, wrong or wrong but reasonable. When using this approach however, it was important to sort the original GPS points through an interspersion layer of the final vegetation map so that any points closer than two pixels from a boundary between classes was filtered out. This restricted the interpreters to judging the classification rather than the accuracy of an artificial boundary.

The initial vegetation maps for the Southern New England Gap Analysis program produced by this procedure had an overall accuracy for all classes of 89.7 % (83.6 % Kappa) for an Anderson level three classification (Anderson et al. 1976). Mis-classifying oak dominant communities as oak/maple/birch co-dominant communities caused the worst user's accuracy of 74 %. The worst producer's accuracy, 80 %, was from the same problem. These results were adequate for the Gap Analysis program's goal of 85 % accuracy and considerably better than the 70 % levels achieved in previous classifications of the same imagery without the sequential rules procedure. However, Maine, New Hampshire and Vermont had collected aerial video coverage for their state Gap Analysis projects at the same time as Southern New England. Those programs developed their own mechanisms for driving their classifications and verifying the results with similar success. Therefore, it would appear that it is the number of ground reference points available for modelling, rather than the specific expert system used to interpret them, that is important for the success of this approach in Landsat classification.

Having established that large-scale sampling with aerial videography was successful in classifying its Landsat data, the National Gap Analysis Program made the technique available to other state gap projects. The Southern New England Gap Analysis program worked in conjunction with

the Gap projects of West Virginia and Colorado to build eight additional Hi8 systems and distribute them as necessary to cover each state that requested one, holding workshops on their use through the National Gap Program and the U.S. Fish and Wildlife Service. Some states such as Maryland and Tennessee used the New England inference rules approach to re-label the unsupervised classifications of their Landsat data, but most developed their own mechanisms to fit the individual needs of their respective projects. In all, 35 states used some form of aerial videography in their projects, either to drive a classification or to verify the results.

A great deal was learned about the advantages and limitations of this system during the Gap Projects. The chief advantage was low cost of operation and ease of use. Because the system could be mounted to the window of any Cessna aircraft, it was easy for projects to find a pilot and aeroplane willing to fly their grids. A GPS tracking program, Geolink Powermap, was used to display both the flight lines and aircraft position in real time during the flight, making it easy for the pilot to follow predetermined tracks.

The biggest problem with the system was preserving the data. Viewing video images on a monitor with a freeze frame mechanism damages the tape, removing thin strips of image and eventually destroying the alignment of the images to the transfer head. Using copies of the tapes for interpretation and archiving the originals could avoid this, but only at a considerable loss of image quality and an understanding that the videotapes will eventually self-destruct even under archive conditions, slowly absorbing water until they become unplayable. Several projects eventually eliminated this problem by digitizing the video data for on-screen interpretation on the computer, but this was a difficult and expensive process with analogue video and most researchers did not have the available disk space to store all their video data as computer files.

Some interpreters also had difficulty visually matching the wide-angle video images and Landsat image to more accurately identify the location of zoomed frame samples. This step was necessary because the GPS in the original set-ups only gave the aircraft's geographic position and did not project the camera's orientation to the location of each frame on the ground. The wide-angle image was at a small enough scale that you could usually identify its surface pattern in the Landsat image, but this slowed down the process of interpretation in non-distinct areas.

The chief criticism of this approach from photographic interpreters and other critics of the Gap Analysis project however was the unavailability of stereo image pairs, which decreased interpretability of the crowns in a forest and limited information on its height and structure, important factors in

evaluating habitat. Stereo could be achieved by grabbing two video frames that overlapped by 60 % and printing them out to view in an optical stereoscope, but that was a time consuming process.

To resolve these limitations, efforts were undertaken to move the aerial videography system into a digital format and improve both its accuracy and interpretability (Slaymaker et al. 1999). The analogue camcorders were replaced with DV (Digital Video) cameras, recording to one-hour DAT tapes that could be reproduced without image quality loss (Figure 18-2). This imagery is also easily transferred to a hard disk in its original compressed format, wrapped in a Quicktime shell for computer access.

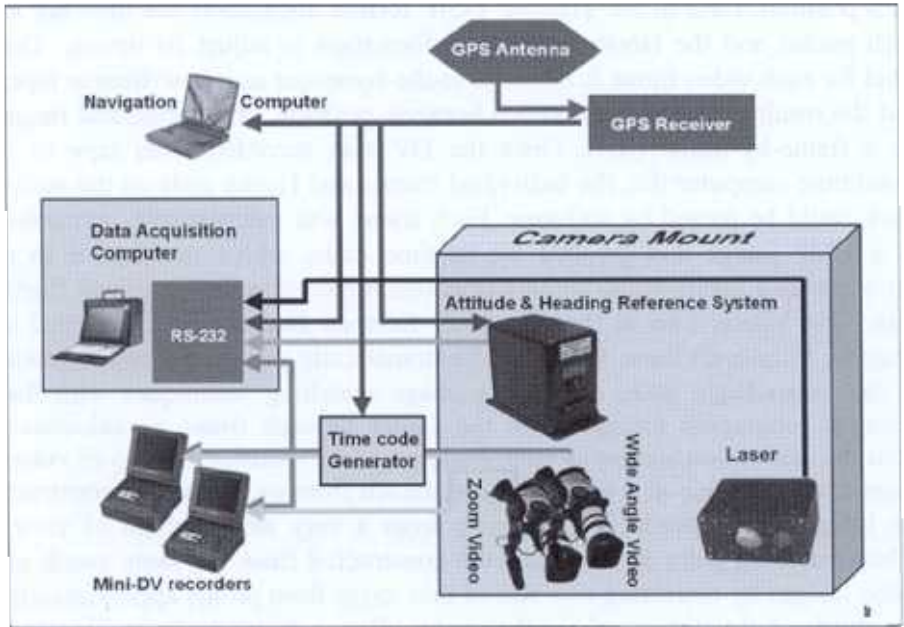
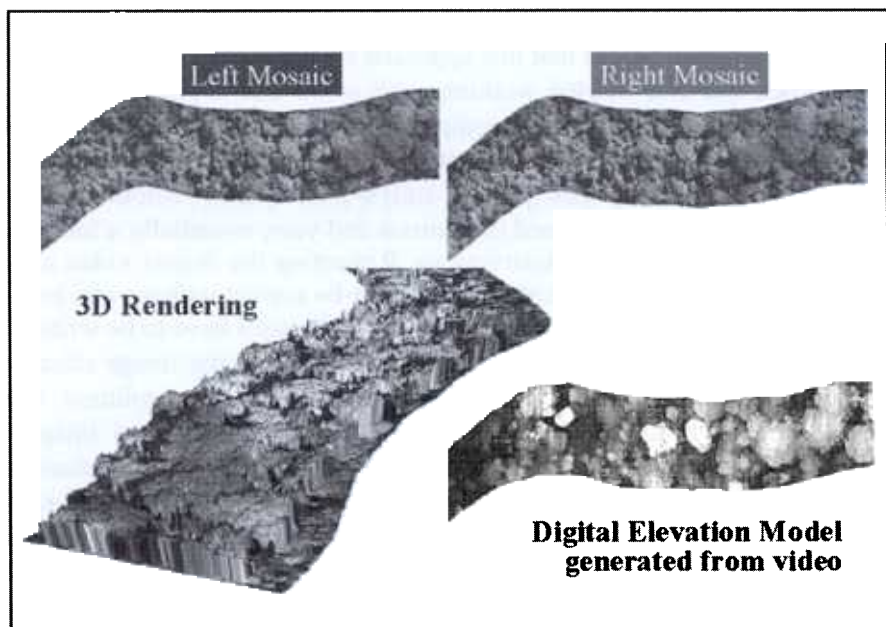


Figure 18-2. Layout of the digital video system.

A Watson attitude and heading reference system was added to the camera platform to determine orientation during flight, recording tip, roll and azimuth eleven times per second, and a pulse laser was added to record a profile of ground and canopy along each flight line. Firing 240 pulses per second, it was set to record the last return of each pulse so as to maximise penetration of the canopy. Horita SMPTE time code and GPS positions were still written to the audio track of the videotape, but the data were also recorded to a separate computer with the Watson and Laser data, using a National Instruments card to time-stamp each entry. The laser and Watson data were initially time-stamped to match the direct NMEA digital output from the GPS receiver. However, it was discovered that most GPS units

have a variable time lag between the receipt of a one-second signal set from satellites and the processing and delivery of the calculated position to a computer. This can introduce an error of up to a quarter of a second between the GPS position and the other data. The Trimble receiver used in this system could be programmed to send a single immediate pulse on receipt of a signal, but this pulse proved difficult to capture and program with the National Instruments card. A simpler solution was not to use the ASCII data output from the receiver directly, but to use the SMPTE data from the Horita GPS3 time code generator instead. The Horita GPS3 unit actually keeps time by counting video frames, labelling each and predicting the next zero frame GPS position. Data in the Trimble T-SIP format documents the time lag in each packet and the Horita uses this information to adjust its timing. The label for each video frame is recorded to the computer as it is written to tape, and the result is an excellent match between position, orientation and range on a frame-by-frame basis. Once the DV was recorded from tape to a Quicktime computer file, the individual frames and Horita code on the audio track could be parsed by software. Each frame was automatically extracted as a BMP image and labelled by its time code, which matched it to a corresponding position and camera orientation record in the combined flight data. The Vision Lab at the Computer Science Department developed a program to convert these frames into automatically geo-referenced mosaics of the video flight path, combining image matching techniques with the accurate geographic placement of the centre of each frame as calculated from the orientation and range data. Because of the extreme overlap of video frames, strips as thin as 4 pixels were extracted from each frame to construct the image swath, producing a mosaic from a very narrow field of view. When matching pairs of mosaics were constructed from the same swath of video images by extracting two sets of thin strips from points approximately two-thirds of the distance from the centre of each frame to its trailing and leading edges, they formed perfect stereo images in a stereoscope or 3D viewing program. These mosaics were actually consistent epipolar models from which a parallax image could be extracted and converted into a digital elevation model with vertical control points from the laser profile (Figure 18-3). This processing essentially mimics the products of more expensive commercial scanning laser and Lidar systems, which produce a direct elevation model of the ground of the ground by scanning across the swath. The difference is that the pulse laser provides a single meter-wide profile, which is then used as vertical control points to transform the parallax model generated from the video strips into a digital elevation model by calibrating it to the ground.

With this system, digital video could be transformed into automatically georeferenced strips of image and superimposed directly over the Landsat image for interpretation. Using the ERDAS program, Stereo Analyst, the epipolar strips could be viewed as stereo images in the same co-ordinate space, with vertical features measured in the image and recorded as text or a georeferenced 3D shape file (ESRI vector format). This approach allowed photographic interpreters to examine the terrain in stereo, outline polygons of vegetation, and then transfer them directly to the underlying Landsat image. They could also measure stand height, relative tree height, or slope and aspect within the images, viewing them directly on screen as optical stereo pairs with either anaglyph red/blue glasses or, more successfully, with Polaroid glasses. The latter block vision to each eye alternately in synchrony with the corresponding right or left image on screen. This takes place at the refresh rate of the monitor, so the viewer is only aware of seeing full colour 3D on screen.



*Figure 18-3.* Image products from the dual mosaic strip construction of epipolar models.

By the time these improvements were available to the Gap Analysis programs, only Georgia, South Carolina and Alabama still needed coverage flown. They have used the new digital 3D system and it remains to be seen if the photographic interpreters for those states find the improvements sufficient to warrant the extra work and cost of obtaining them.

The DV camera set-up has also been used in other applications beyond Gap Analysis, flying approximately 170 hours of video across the Amazon Basin to provide ground verification for cover mapping with JERS-1 mosaics by the NASA LBA-Ecology Project's Global Rain Forest Mapping program (GRFM). It compared canopy height and forest structure to Synthetic radar (SAR) sensor response (Hess et al. 2002). In 2001, development of the system was taken over by Winrock International, a non-profit research foundation devoted to the development of new techniques for sustainable agriculture and forest management. Winrock is currently working with The Nature Conservancy on a joint grant from the Department of Energy (DOE) to develop new methods of estimating the carbon sequestered in tropical and temperate forests. The foundation is exploring the use of 3D videography to extrapolate ground site estimates of standing biomass across a forest region by developing positive correlations between crown diameter / height and dbh (diameter at breast height). Previous work with Winrock and NASA-LBA (Slaymaker et al. 1999; Hayward and Slaymaker 2001) indicates that this approach should work.

Winrock has also started working with some true digital cameras that record their data directly to a computer in an RGB or RGBIR format. Although a considerable improvement over analogue, DV still suffers from low resolution (480 by 720 pixels) and a poor quality colour-recording model (NTSC) that is composed of contrast and yaw, essentially a black and white image with colouring instructions. Recording the digital video signal to DAT tape format means that the tapes can be copied without any loss of image quality, but it also means that the original images have to be written in a lossy compression mode that has a definite effect in the image structure. By contrast, there are now relatively inexpensive high resolution RGB digital cameras available that produce a 2,000 by 3,000 pixel image of higher quality than a scanned 35 mm negative and reverse the economic limitations of the 35 mm format as compared to aerial video. Working with the Norwegian Space Centre, Winrock flew a Gap Analysis style survey of parts of that country in the summer of 2001, using a Kodak DSC 760 RGB digital camera, and exposing 3,000 to 5,000 frames per flight that were saved to a hard disk as 7 megabyte georeferenced tiff images. A firewire link between the camera and computer controlled all camera adjustments except focus and enabled the downloading of an image to disk at a maximum rate of one exposure every five seconds. An on-screen preview allowed the operator to compare each image to the previous exposure and determine if there was sufficient overlap. The camera software has a built-in intervalometer that can be set to fire the camera on the second at a specific interval. That timing is based on the computer's internal clock rather than GPS time, but one of the

Norwegian programmers was able to write a simple program that synced the computer's clock to incoming NMEA GPS data. The camera records the time of each exposure to 1/1000<sup>th</sup> of a second by computer clock time, so exposure positions between GPS records can be calculated. It can also input NMEA GPS data through its own serial port and write it directly to the header of each exposure if the last GPS record is sufficient. The Kodak DCS 760 costs around \$6,500 and can be operated at full speed by any recent laptop with built in firewire support and at least 520 megabytes of ram. This means that an effective high-resolution digital aerial system can be assembled for less than \$10,000. Kodak has also developed a 4,080 by 4,080 pixel camera back that attaches to the Mamiya AFD or Hasselblad medium format cameras and offers most of the same remote control capacities as the DCS 760 (although it cannot download GPS data directly). It costs around \$12,000 for the back alone, in addition to the cost of the front-end camera with a wide-angle lens. This back can be operated with the same type of laptop as the DCS 760, but with a significantly longer download time. The maximum firing rate of these digital cameras is a significant issue in determining their suitability for forest surveys and frame sampling applications. The minimum interval of 12 to 15 seconds between exposures required by the Kodak 4,080 by 4,080 pixel digital camera back limits its aerial exposures to scales of 0.7 meter per pixel or more for stereo coverage (60 % overlap), restricting its use to large area coverage at higher altitudes. The smaller image size and rectangular format of the Kodak DCS 760 allows it to fire stereo coverage at scales down to 0.5 meter per pixel with a 3,000-pixel swath, or 0.3 meter per pixel when turned sideways for a 2,000-pixel swath. This is the format that was used in the Norwegian Gap Survey.

The other major limitation of this camera and most high-resolution digital cameras on the commercial market is that they use a matrix filtering system to construct colour images from a single monochrome CCD sensor. Each pixel is coated with a different colour filter in a pattern of 25 % Red, 50 % Green, and 25 % Blue. The images are saved as a single layer image from which three layer RGB images are generated in software by the extrapolation of colour information to each pixel from its neighbours. The advantage of this approach is that the raw images are much smaller than the final product, making it easier to transfer and store very high-resolution files on a laptop with limited hard disk space. A 2,000 by 3,000 12-bit image can be captured and stored as a 7-megabyte file, then processed to a 35-megabyte 32-bit image after the flight. The disadvantage is a loss of some radiometric information and colour resolution as compared to a true RGB digital camera that uses 3 CCD monochrome chips behind colour filters.

High-resolution versions of 3 CCD cameras are still relatively rare and expensive, but Winrock is also working with a 3 CCD RGBIR multi-spectral digital camera, the Duncantech M4100, with a resolution of 1,024 by 1,920 pixels. The camera uses a 3-way prism behind the lens with two monochrome CCDs, recording red and near infrared images matched to Landsat bands 3 and 4, plus a colour matrix chip recording the green and blue bands. Since the colour chip is 50 % green, there is minimal resolution loss in the extrapolation of that band and while the blue colour layer is still generated from 25 % of the chip, but it is the lowest resolution band used in aerial imagery anyway. The output is processed to a four band RGBIR image in the camera, and the gain of individual bands can be adjusted in flight, allowing it to be calibrated to known spectral responses with a Barnes radiometer and ground panels. At \$17,000.00, the M4100 is a fraction of the cost of most multi-spectral digital systems, which start around \$125,000 as multi-camera arrays. It is connected to its host computer by a framegrabber or direct camera link, rather than firewire, allowing it to record images up to 10 frames a second at resolutions of less than 10 centimetres per pixel and overlaps of 80 % or more. This means that the Duncantech can replace digital video in large-scale applications that require high overlap if the operator has a computer that can handle that rate of data transfer. Even firing the Duncantech at a more modest rate of 5 frames a second (7.5 megabytes per image) accumulates data at a rate of 130 gigabytes an hour. Recent increases in computer speed and hard drive size make this rate of data acquisition feasible. Winrock's field computer (a lunchbox luggable) holds five 73-gigabyte 160 LVD SCSI drives, stripped to a single 266-gigabyte fault-tolerant array, which limits actual data collection in flight to about 2 hours. That data has to be copied to portable 160-gigabyte firewire drives after each flight, then written to Exabyte tapes over the next several days.

How this approach will work with large-scale sampling projects like the National Gap Program or the NASA-LBA coverage of the Amazon Basin remains to be seen. Most of the Gap Analysis states and the Amazon Basin were flown by starting in one corner of the grid and following the grid pattern until done, landing at local airports and staying at motels along the way. Organising the necessary data storage and transfer with that kind of schedule would be difficult now, but should become easier as larger IDE drives (200 to 500 GB) become available in firewire enclosures and their prices continue to drop. At this time, it costs about \$1.22 per gigabyte to store image data on Maxtor 5400 rpm 160 GB hard drives, which is less than the cost of storing the same data on Exabyte tape. This trend of faster computers with larger hard drives should continue in the future, making the handling of these large data files more practical.

The Computer Science Department at the University of Massachusetts is also working on a program that will write the stereo mosaics in real time during the flight, discarding the bulk of each image. This would solve the storage problem, but will also require faster computers than are available today, as well as a considerable leap of faith on the part of the camera operator.

### 3. CONCLUSION

Experienced aerial photographic interpreters have a heuristic ability to identify different tree species and plant communities at large scale under a range of slopes and lighting conditions. This is a uniquely human trait that has not been rivalled by the automated analysis of spectral signatures from satellite data. The Gap Analysis Program developed a practical frame sampling method that utilised this human capacity to improve the machine classification of Landsat data over large regional areas. The efficacy of this multi-scale approach to interpreting satellite imagery has proven itself over the life of the project.

During the 12 years of this effort, there was a general trend toward improving the quality and accuracy of the data with a corresponding increase in the cost and complexity of the equipment. The first dual camera Hi8 system cost approximately \$7,000. Switching to the DV system with an attitude indicator and laser increased its cost to \$32,000, while the complete Duncantech system runs over \$60,000 (including the on-board computing system and the several terabytes of data storage needed to manage the imagery post-flight). Analogue video systems are practically extinct and the cost of DV camcorders has dropped considerably, so it would still be possible to put together a simple video system for \$7,000 that had the advantage of digital data storage on DAT tape. The major expenses in building more complex systems lie in either capturing the orientation of the cameras at the moment of each exposure for automatic georeferencing / mosaicking programs, or moving into higher resolution, multi-spectral imagery. The researcher or forest manager who can meet his needs with natural colour imagery that has a more approximate geographic position attached to it can still put together an excellent imaging system, consisting of either DV or digital still cameras, that clamps to a Cessna and costs under \$10,000. When the Gap Analysis Program initiated its aerial frame sampling system, the driving factors to choose video over the higher resolution of scanned 35 mm film were the per frame costs, automatic operation of the camera outside the aircraft without having to change film and the ability to

tag each video image with a specific geographic position between the once per second GPS signals. With the evolution of inexpensive high-resolution digital cameras in 35 mm camera bodies, those advantages have been essentially eliminated. The digital frame cameras have better colour and resolution than digital video and bypass the compressed tape storage that limits the quality of DV images. The major limitation of these digital cameras is the long time interval between exposures, which is required to download each image through the relatively inexpensive firewire links. Cameras like the Duncantech or the Atmel (2,300 by 3,500 pixel) digital camera use more expensive framegrabbers or direct camera links and can download at much higher frame rates, but with a corresponding increase in cost and data management problems.

All of these systems are still very inexpensive, however, compared to commercial multi-spectral or hyper-spectral digital cameras, or scanning laser/video and Lidar systems, most of which will do a better job of aerial data collection, albeit for considerably more money. The history of video and digital aerial camera systems, like that of small format aerial photography in general, has been the development of "jury-rigged" systems to meet specific resource management objectives within severe budget limitations. What they lack in photogrammetric accuracy and spatial resolution is compensated for by their ease of mobilisation and subsequent temporal resolution: the ability to be available and affordable as needed for resource monitoring.

## REFERENCES

- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. *U.S. Geological Survey, Professional Paper 964*. Washington, D.C.
- Drake, S. (1996). Visual Interpretation of Vegetation Classes from Airborne Videography: An Evaluation of Observer Proficiency with Minimal Training. *Photogrammetric Engineering & Remote Sensing*, 62, 969-978.
- Dunford, C., Mouat, D., Norton-Griffiths, M., & Slaymaker, D. M. (1983). Remote sensing for rural development planning in Africa. *The Journal for the International Institute for Aerial Survey and Earth Sciences*, 2, 99-108.
- Eleveld, M. A., Blok, S. T., & Bakx, J. P. G. (2000). Deriving relief of a coastal landscape with aerial video data. *International Journal of Remote Sensing*, 21, 189-195.
- Graham, L. A. (1993). Airborne Video for Near-Real-Time Vegetation Mapping. *Journal of Forestry*, 8, 28-32.
- Hayward, C. D., & Slaymaker, D. M. (2001). Estimating the significant above ground biomass of Amazonian rain forest using low altitude aerial videography. *18<sup>th</sup> Biennial*

- Workshop on Color Photography and Videography in Resource Assessment, in press. 16-18 May, ASPRS, Bethesda, Maryland.
- Hess, L. L., Novo, E. M. L. M., Slaymaker, D. M., Holt, J. Steffen, C., Valeriano, D. M., Mertes, L. A. K., Krug, T., Melack, J. M., Castil, M., Holmes, C., & Hayward, C. (2001). Geocoded Digital Videography for Validation of Land Cover in the Amazon Basin. *Journal of Remote Sensing*, in press.
- Jennings, C. A., Vohs, P. A., & Dewey, M. R. (1992). Classification of a wetland area along the upper Mississippi River with aerial videography. *Wetlands*, 12, 163-170.
- Kelly, P. M., & White, J. M. (1994). *Preprocessing remotely-sensed data for efficient analysis and classification*. Computer Research Group, MS B-265. Los Alamos National Laboratory.
- King, D. J. (1995). Airborne multispectral digital camera and video sensors: a critical review of system designs and applications. *Canadian Journal of Remote Sensing*, 21, 245-273.
- Marsh, S. E., Walsh, J. L., & Sobrevila, C. (1994). Evaluation of airborne video data for land-cover classification accuracy assessment in an isolated Brazilian forest. *Remote Sensing of Environment*, 48.
- Mausel, P. W., Everitt, J. H., Escobar, D. E., & King, D. J. (1992). Airborne videography: current status and future perspectives. *Photogrammetric Engineering and Remote Sensing*, 58, 1189-1195.
- Meisner, D. E. (1986). Fundamentals of airborne video remote sensing. *Remote Sensing of Environment*, 17, 63-79.
- Norton-Griffiths, M. (1988). Aerial point sampling for land use surveys. *Journal of Biogeography*, 15, 149-156.
- Richards, J. A., Landgrebe, D. A., & Swain, P. H. (1982). A means for utilizing ancillary information in multispectral classification. *Remote Sensing of Environment*, 12, 463-477.
- Sayn-Wittgenstein, L. (1978). Recognition of tree species on aerial photographs. *Information Report FMR-X-118*, Forest Management Institute, Canadian Forestry Service, Ottawa, Ontario.
- Scott, J. M., & Jennings, M. D. (1998). Large-Area Mapping of Biodiversity. *Annals of the Missouri Botanical Garden*, 85, 34-47.
- Sidle, J. G., & Ziewitz, J. W. (1990). Use of aerial videography in wildlife habitat studies. *Wildlife Society Bulletin*, 18, 56-62.
- Slaymaker, D. M., & Hannah, L. (1997). GPS-logged Aerial Video as a Georeferencing Tool For Digital Imagery in Remote Regions, a case study in Madagascar. *The Proceedings of the 16th Biennial workshop on Color Photography and Videography in Resource Assessment*. ASPRS, Bethesda, Maryland.
- Slaymaker, D. M., Jones, K. M. L., Griffin, C. R., & Finn, J. T. (1996). Mapping deciduous forests in southern New England using aerial videography and hyperclustered multi-temporal Landsat TM imagery. Scott, J. M., Tear, T. H., & Davis F. W. (Eds.). *Gap Analysis: A Landscape Approach to Biodiversity Planning*, 87-101. American Society of Photogrammetry and Remote Sensing, Bethesda, MD.
- Slaymaker, D. M., Schultz, H., Hanson, A., Riseman, E., Holmes, C., Powell, M., & Delaney, M. (1999). Calculating Forest Biomass with Small Format Aerial Photography, Videography, and a Profiling Laser. *The Proceedings of the 17th. Biennial workshop on Color Photography and Videography in Resource Assessment*, 241-260. April 1999, ASPRS, Bethesda, Maryland.
- Strong, L. L., & Cowardin, L. M. (1995). Improving prairie pond counts with aerial video and global positioning systems. *Journal of Wildlife Management*, 59, 708-719.

- Stumpf, K. A., & Koltun, J. M. (1993). Rule based aggregation of raster image classifications into vector GIS databases with five and forty acre minimum size-mapping units. *ASPRS/ACSM Conference*, 261-278. Washington, D.C.