

SAS-Implementations Supporting Satellite, Aircraft, and Drone-based Remote Sensing Endeavors and Their Influences in the Classroom

Cecil Hallum

Sam Houston State University, Huntsville, TX

ABSTRACT

SAS has been a significant “partner” over a career spanning 40 years for this researcher. This presentation summarizes key SAS applications in satellite, aircraft and drone-based remote sensing endeavors (beginning in the early ‘70’s when SAS was first implemented at NASA/Johnson Space Center). The coverage includes a discussion of recent multivariate strategies implemented in SAS geared toward finding missing bodies in digital imagery collected from drone flights as well as current research oriented toward improving the accuracies and speeds of such capabilities. Discussion of the impact of this research in the classroom for educational purposes at the university and high school levels is emphasized as well. This presentation is also being done as a follow-up to a request from a SAS representative in attendance at a 2007 SCSUG meeting. Moreover, an in-depth coverage of these topics (particularly the analytical and SAS programming specifics) will appear in the practicum of an SHSU graduate student in May 2012.

INTRODUCTION

This author has been involved with remote sensing, GIS, and image analysis over a key part of his career beginning as long ago as in the mid-70's and early 80's at NASA/Johnson Space Center and, since that time, in numerous other areas as well. As a result, a number of remote sensing efforts have been conducted (with some remaining in progress) at Sam Houston State University with each having a direct dependence on the numerous capabilities of SAS. These efforts are listed summarily below.

1. **Wheat Area and Production Estimation:** At NASA/Johnson Space Center, remote sensing was instrumental, beginning in the early 1970’s, in support of estimation of wheat area and production throughout the world (these efforts were directly a result of the largest remote sensing endeavor, the Large Area Crop Inventory Experiment (LACIE), that has been conducted to-date; it began back in the early 1970’s and was headquartered at Johnson Space Center in Houston, Texas. The author headed up the sampling and estimation component for the LACIE in support of global wheat area estimation using the Landsat satellite technology.

2. **Monitoring Landscape Changes:** A second key research effort participated in by the author was in support of the Department of Defense (DOD) as a result of it receiving a mandate from the Environmental Protection Agency (EPA) to monitor “changes to the landscape” at military training facilities. In particular, the EPA indicated to DOD that the demand for land at military bases for the purpose of practicing military maneuvers, including bombings and other destructive missions (destruction of vegetation from tanks and other land vehicles), resulted in exceeding the supply of land available for training. In support of this DOD effort, SHSU researchers initially developed a “change assessment” capability (see Hallum (1993), Hallum (1999) comprising 4 technical reports in 1999, and Hallum (2000)) using satellite and aircraft data as inputs to a multivariate statistical strategy that permitted the identification of where changes had occurred but also permitted quantifications of the “severity” of the change at the pixel level. The knowledge gained from this effort led to a number of additional activities including another more recent, and especially worthwhile and rewarding, endeavor described below.

3. **Miniaturized Remote Sensing:** Recent emphases (over the past 6 years) resulted in research, development and testing that has involved a number of students from our M.S. program (Minter (2006), He (2007) and Demel (2008)) in an area that might be described as: “affordable, miniaturized remote sensing”. Here the goal was to utilize very affordable (off-the-shelf) vehicles, unlike the very expensive operations used by the military, and, consequently put in place a viable and very cost-effective capability for doing precisely what R. P. Flight Systems in Wimberly, Texas (<http://www.rpflightsystems.com/MainPage.html>) and EquuSearch (located in Webster, Texas) (<http://texasequusearch.org/>) have as one of their key goals: locate missing bodies (or objects of any kind) using digital photos taken by digital cameras (or other type sensors) aboard a remote control drone.

4. **Spinoff Remote Sensing Projects:** A number of additional efforts received research emphasis recently at SHSU as a part of the graduate M.S. Program in Mathematical Statistics. These have included:
 - a. Research efforts (Hallum (2010)) to support employing sound as another dimension for analysts’ insight into assessing key information in digital photographs as well as images acquired by satellites, aircraft, and drones.

- b. Research (Minter (2006) and Hallum (2010)) to assist the seeing impaired to “hear” information in digital photographs (i.e., whereby “hearing” is used as another “dimension” for retrieving information from imagery (whether it be from digital photography, from satellite images, aircraft data, or from other sources)).

In summary, the purpose herein is to provide details that have transpired over the past 40 years of SAS applications in support of the following sequence of research agendas:

LARGE SCALE (I.E., GLOBAL) REMOTE SENSING



LOCALIZED LEVELS OF REMOTE SENSING (I.E., MONITORING LOCAL CHANGES DUE TO EARTHQUAKES, TORNADOES, TSUNAMIS, ETC.)



MINIATURIZED REMOTE SENSING APPLICATIONS (USING OFF-THE-SHELF, COST-EFFECTIVE, REMOTE CONTROL AIRBORNE VEHICLES)

These will be discussed in more detail at conference time and will be detailed in extensive detail in the previously referenced SHSU graduate student’s practicum to be completed in May 2012.

PROJECTS’ CHARACTERISTICS INCLUDING THEIR DEPENDENCIES ON SAS

Wheat Area and Production Estimation:

Several decades of research in the area of aerospace remote sensing has resulted in technologies that provide the means to economically provide better crop forecasts (see Hallum et. al. papers dating from 1972 through 1994). In 1974, the Large Area Crop Inventory Experiment (LACIE), a joint effort of the National Aeronautics and Space Administration (NASA), the United States Department of Agriculture (USDA), and the National Oceanic and Atmospheric Administration (NOAA), began applying this technology on an experimental basis to forecasting harvests in important wheat-production areas. In 1976, the author became a NASA employee and a member of this team consisting of approximately 40 independent university, industry, and government scientists and researchers working through the directions of Johnson Space Center.

Landsat satellites have been used for several decades in a wide variety of applications, including climate research, natural resources management, commercial and municipal land development, public safety, homeland security and natural disaster management. Figure 1 below provides a schematic of the lifetimes of the 7 U.S. landsat satellites. Notice that, except for Landsats 5 and 7, the Landsat satellites are almost depleted; plans are in motion to launch another in 2012: the LDCM (Landsat Data Continuity Mission) satellite.

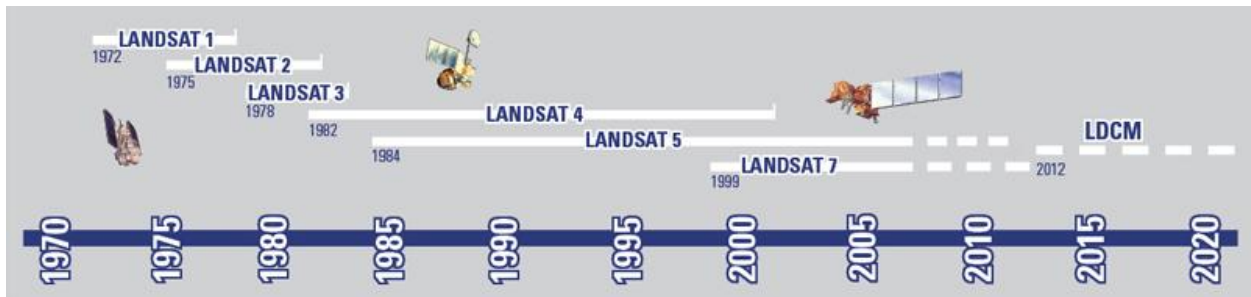


Figure 1: Chronology of Landsat Satellites

Since 1863, crop surveys in the Department of Agriculture have expanded greatly until today a large volume of agricultural estimates is published on a periodic basis. The substantial expansion in the volume of agricultural data has not been paralleled by major improvements in estimation methods, which is somewhat distressing in view of the significant developments in the theory of sample designs – particularly in the past 60 years. Until efforts of the Statistical Reporting Service (SRS), now renamed to the National Agricultural Statistical Service (NASS), and the Large Area Crop Inventory Experiment (LACIE), the typical procedure has been mailed inquiries for collecting basic data and an assortment of techniques for removing bias in the transformation of raw data into published estimates. Incidentally, notice in Figure 1 that the Landsat 1 satellite was launched in 1972. That year the author was employed at Lockheed in Houston and was the Chair of the “Signature Extension Team” of the LACIE effort; this team had the duty of determining the extent to which the Landsat 1 data could be used to distinguish between differing landscape canopies sensed by its multi-spectral sensor. The initial case was one of verifying whether one could distinguish between land and water using the satellite data. On the initial flight of Landsat 1 over Lake Livingston, the Signature Extension Team members identified several key spots on the Lake (actually 8 carefully selected locations out on the Lake body) to observe water conditions in support of assessing problems (if any) in accomplishing this task. The distinction between land and water fortunately turned out to be an easy case and the rest of LACIE is well-documented in the

Proceedings of the Large Area Crop Inventory Experiment (LACIE) Symposium (Vol. 1 and 2), held at NASA/JSC, over the period: October 23-26, 1978.

In the early '70's, the author requested SAS to be implemented at NASA/Johnson Space Center. This request was approved and, consequently, SAS was a significant part of the development of the capabilities described above and, more specifically, in the areas listed below (and in other parts of the space program not listed below):

- **Configuration and Geographical Extent of the Sampling Unit:** The sampling unit eventually used in the LACIE was a 5- by 6-nautical-mile rectangle. Two of the most important considerations leading to this choice for sampling unit size were:
 - This size (even though somewhat large) proved to be satisfactory in terms of serving that analyst's needs and permitting required sampling precision without creating an unmanageable data load (see Perry and Hallum (1979) for a more detailed analytical look into the sampling unit issue).
 - Also, this size was sufficiently large to provide the analyst with a good perspective on the variety and distribution of crops within a given locality.
- **Sample Selection Procedure:** SAS supported initial testing of the effectiveness that could be expected from the use of a stratified sample procedure; from a Neyman allocation (see Feiveson and Hallum (1972) and Hallum and Basu (1972)) that relied on a "new sampling strategy"; in particular, the new sampling strategy was one that relied on strata that were geared toward being more homogeneous in regard to agricultural practices, field sizes, and similar characteristics. One key end result was a reduction of the required number of sampling units from over 1900 units down to slightly over 600. The Neyman allocation formula was implemented and exercised in SAS. Nowadays, PROC SURVEYSELECT and PROC SURVEYMEANS would do a key part of this effort with considerably more ease; these procedures will be used as a key part of the graduate sampling course to be taught in the spring 2012 semester at SHSU.
- **Allocation of Sampling Units:**

The allocation of sampling units was, again, based on the use of a stratified random sampling procedure; again, the Neyman allocation approach (see Cochran (1977)) was the allocation approach applied (and a part of the somewhat new PROC SURVEYSELECT).

- **Area Estimation and Classification Approach:** The Landsat data comprises as much as 7 dimensional data and oftentimes the optimal classification scheme was that of the maximum likelihood approach under the multivariate normal assumption (see Johnson & Wichern (2002)) SAS was utilized a priori to do some initial testing to see how well this approach would work. Minter and Hallum (1973) developed and tested a “thresholding” procedure for classifying pixels into various categories of vegetation. The technique depicted in Figure 2 (to be explained at conference time) reduced the classification time significantly.

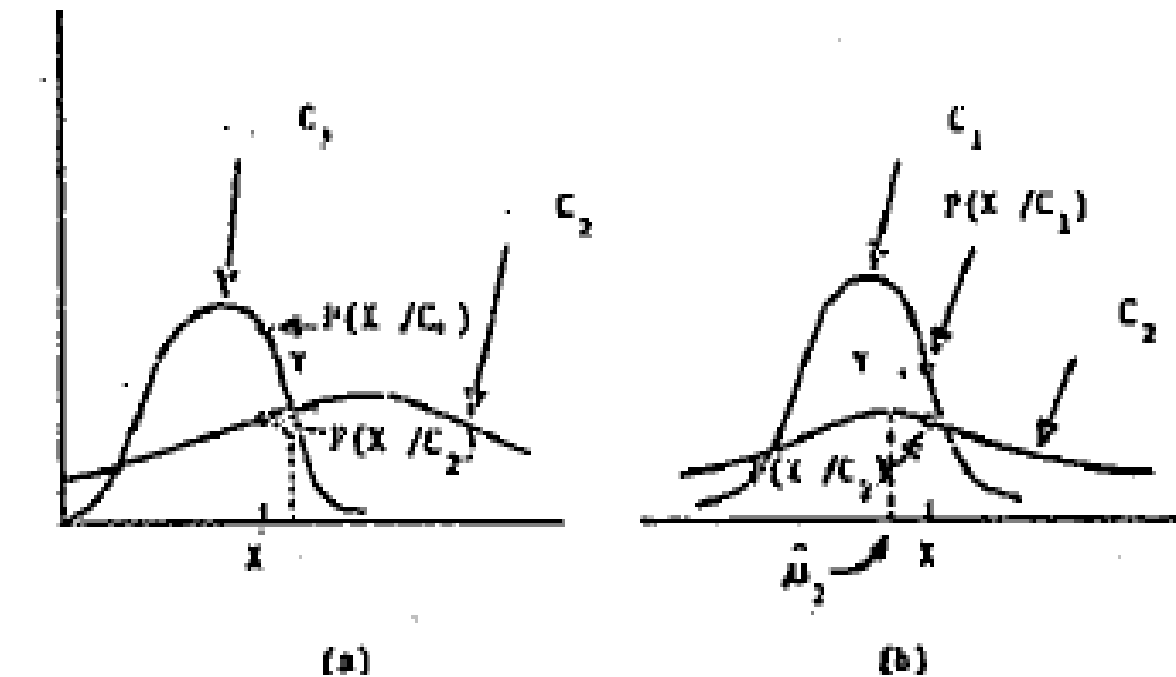


Figure 2: Strategy to Accelerate the Process of Maximum Likelihood Classification

Briefly, note in Figure 2, details will be provided for how the γ is found and how its usage can save considerable computing time to permit speeding up the maximum likelihood classification procedure (a multivariate scheme for classification of unknowns into one of several classes of things).

- **Variance Estimation:**

In carrying out variance estimation, this researcher along with numerous others utilized SAS to play the “what if” game, to test underlying theories, and to modify and refine textbook approaches where needed. This required numerous simulations and considerable testing which was readily

available in SAS at that time. Now SAS has added considerably more capabilities along the lines regarding what was needed then with the addition of PROC SURVEYSELECT, PROC SURVEYFREQ, PROC SURVEYMEANS, PROC SURVEYLOGISTIC and PROC SURVEYREG. Each of the various procedures during the LACIE were mandated to be written as a dedicated piece of software using the programming language of preference at that time (e.g., the C language, Fortran, ADA, etc....I do not recall), consequently, SAS was not a part of the operational software. Again, SAS was used in a research mode during the LACIE.

Miniaturized Remote Sensing: The methodology development at SHSU has already led to an initial capability (see Demel (2008)) whereby several thousand digital pictures are digitized, statistically analyzed (using multivariate and related techniques), followed by a ranking (i.e., ordering) whereby an image analyst would, expectedly, find the missing entity (e.g., a missing body or any object being sought) within the first few photos (e.g., in the first 20 or 30) and, thereby, eliminate the need to pore over the 2 to 3 thousand photos typically taken in one drone mission. Efforts have also been conducted regarding providing critically needed information following the aftermath of crises such as that experienced in the Haiti and Chili earthquakes, hurricanes, tsunamis, etc. An example of the later is that of investigating ways and means of locating the “tent cities” (i.e., locations of the homeless) along with estimates of the number of homeless persons. A practicum was completed somewhat recently by an SHSU statistics graduate student (see He (2010)) related to this particular topic.

The initial statistical approach discussed herein and implemented in SAS in a GUI framework (see the Appendix 3) is described in the following steps:

1. Each digital photo is converted into pixel-level digital data, composed of a series of vectors of length 3, containing each pixel’s red, green, and blue (RGB) values. This process is done using the “screen control language (SCL)” component of the Applications Frame development module in SAS as part of a GUI (graphical user interface) which increases a user’s ease when interfacing with other needed SAS modules.
2. Augmentation of each pixel vector, to include spatial information more readily; this method is a modification of the technique suggested by Cressie (1993).
3. Cluster analysis is used to simplify targeted areas within a digitized image and assists in expediting the process of “detecting” a target object. Once the clusters that make up the target object are known (by referencing objects that “look” similar to the target) one can restrict the

search to only those areas that are made up of the same clusters as the target itself (further expediting the process).

4. A “moving rectangular window” is employed (relying on the generalized Mahalanobis distance function (see Johnson and Wichern (2002)) to rank “target-like” objects across and within images. Based on these rankings, the time required for analyzing a larger assortment of images (which an image analyst currently has to do manually) has been reduced considerably (see Demel(2008)).

The specific logic resorted to is detailed further below:

1. Let $K=\{r, s, t\}$ comprise the cluster IDs of the three clusters comprising the largest proportion of the target (the optimal number of cluster IDs to utilize is awaiting further research).
2. Center the “detection window” at every occurrence of a pixel whose cluster ID is in the set K and compute the following:

$$d_{ij}^2 = (\bar{X}_0 - \bar{X}_{wij}) S_{pooled}^g (\bar{X}_0 - \bar{X}_{wij})'$$

where

\bar{X}_0 = mean vector of RGB values for the target object restricted to only pixels from clusters with IDs from K (of course, for the object itself, this only has to be done once),

\bar{X}_{wij} = mean vector of RGB values from the detection window restricted to only pixels from clusters with IDs in K .

S_{pooled}^g = the generalized inverse of the pooled variance-covariance matrix (again restricted to the same clusters as indicated above). Here the generalized inverse is utilized in case there are segments of a digitized photo that are highly discrete (i.e., for which the inverse does not exist).

3. For the photo, assign the value d_m^2 where:

$$d_m^2 = \min_{ij} d_{ij}^2$$

4. Rank every photo accordingly

$$d_{c_1}^2 \leq d_{c_2}^2 \leq \dots \leq d_{c_n}^2$$

The image analyst can then investigate the photos in the order of the above ascending Mahalanobis distances. Expectedly, the object will be found in the first few (e.g., the first 10 or so) rather than

having to pore over the entire collection (e.g., 2000 or more --- which is currently the case in many situations where drones are used to take photos in the search of an object (e.g., a missing body) – the number of photos over a 2 mile radius oftentimes is as many as a couple thousand). Photos (taken by an 8 megapixel camera) of actual missing bodies is included in Appendix 4.

SUMMARY

Considerable additional specifics will be provided at conference time, however, again a primary purpose of this presentation is to provide a summary documentation of 40 years of SAS usage in support of the following goals:

1. Remote sensing endeavors utilizing satellite and aircraft data (beginning back in the 1970's at NASA/Johnson Space Center).
2. Development, testing and implementation of capabilities to monitor the landscape using data from any source (again including satellite-, aircraft- and also drone-acquired, remotely-sensed data).
3. To provide key insight into how to use off-the-shelf, cost-effective and easy-to-use capabilities for airborne monitoring of localized areas.
4. Other side applications including that of assisting a seeing-impaired individual “hear” information contained in digitized imagery (from any source: satellites, aircraft, drones and/or home photography from the use of any digital camera).

Further details are available via the Web links provided in the first link below; the second link below is to the PowerPoint slides used in the conference presentation:

Link 1: http://www.shsu.edu/~mth_crh/SCSUG/RS_Links.docx

Link 2: [http://www.shsu.edu/~mth_crh/SCSUG/SCSUG-2011_Presentation\[1\].pptx](http://www.shsu.edu/~mth_crh/SCSUG/SCSUG-2011_Presentation[1].pptx)

REFERENCES

Chen, Xiadong and Hallum, Cecil, “Improved Classification Using Weighted Distance Functions”, *Proceedings of the American Statistical Association Conference*, Dallas, Texas, August 1998, pp. 118-123.

Chhikara, Raj, Hallum, Cecil, and Lycthuan-Lee, T.G., “Stratification and Sample Design for Landsat Crop Surveys”, Special Issue of *Communications in Statistics, Theory and Methods*, Vol. 13, No. 23, 1984, pp. 2905-2935.

Cochran, William G., *Sampling Techniques*, 1977, John Wiley & Sons, Inc., New York, N.Y.

Cressie, Noel A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Demel, Seth, *Object Detection, Sense and Avoid, and Near Real-Time Multivariate Analysis Onboard an Inexpensive UAV*, an M.S. thesis completed at Sam Houston State University, August 2008.

Feiveson, A.H. and Hallum, C.R., “The Sampling Design for LACIE”, *Proceedings of the Large Area Crop Inventory Experiment (LACIE) Symposium*, held at NASA/JSC, October 23-26, 1978, Houston, Texas.

Hallum, C.R., “On a Model for Optimal Proportions Estimates for Category Mixtures”, *Proceedings of the Eighth International Symposium on Remote Sensing of the Environment*. The University of Michigan, Ann Arbor, Michigan, October 1972.

Hallum, C.R., “Null Class Thresholding in Gaussian Maximum Likelihood Classification”. Technical Report LEC/HASD No. 640-TR-133 under NASA Contract NAS 9-12200, August 1972.

Hallum, C.R., “Feature Selection via an Upper Bound (to any Degree Tightness) on Probability of Misclassification”, *Proceedings of the Conference on Machine Processing of Remotely Sensed Data*, Purdue University, West Lafayette, Indiana.

Hallum, C.R., “Experiment Design Overview”, *Proceedings of the Large Area Crop Inventory Experiment (LACIE) Symposium*, held at NASA/JSC, October 23-26, 1978, Houston, Texas.

Hallum, C.R., Feiveson, A.H., Chhikara, R., and Houston, A.G., “Sampling and Aggregation in LACIE”, *Proceedings of the Large Area Crop Inventory Experiment (LACIE) Symposium*, held at NASA/JSC, October 23-26, 1978, Houston, Texas.

Hallum, C.R. and Basu, J.P., “A Natural Sampling Strategy”, *Proceedings of the Large Area Crop Inventory Experiment (LACIE) Symposium*, held at NASA/JSC, October 23-26, 1978, Houston, Texas.

Hallum, C.R. and Perry, C.R., “Sampling Unit Size Considerations for Reducing Data Loads in Crop Inventorying Using Remotely-Sensed, Satellite-based Data”, *Proceedings of the IEEE Telecommunications Conference*, Galveston, Texas, November 2, 1982.

Hallum, Cecil R. and Perry, Charles R., “Estimating Optimal Sampling Unit Sizes for Satellite Surveys”, *Remote Sensing of Environment*, January 1984.

Hallum, Cecil R., “A Change Detection Strategy for Monitoring Vegetative and Land-use Cover Types Using Remotely-sensed, Satellite-based Data”, *Jour. of Remote Sensing of the Environment*, Vol. 43, No. 2, 1993, pp. 171-178.

Hallum, Cecil, “Accelerating Maximum Likelihood Classification: An Approach with SAS/IML Software”, *Proceedings of the 19th Annual SAS Users Group Conference*, April 10-13, 1994, Dallas, Texas.

Hallum, Cecil, “Summary of Findings from Change Analysis of the Depth 10 and Depth 20 Surface Water Data”, prepared for contract personnel at the Watershed Experiment Station (WES) in Vicksburg, Mississippi, March 1999; this technical report was in support of my work as a Senior Research Statistician on the TRIES grant with WES.

Hallum, Cecil, “May 1999 Report of Initial Approaches for a Change Assessment Methodology and a Verification and Sensitivity Analysis Capability for Land Management Models”, a technical

report in support of the TRIES grant with the Water Experiment Station (Vicksburg, Mississippi), May 1999 (30 pages).

Hallum, Cecil, “Design of a Change Assessment Methodology and a Verification and Sensitivity Analysis Capability for Land Management Models”, end of contract year technical report in support of TRIES grant with the Water Experiment Station, Vicksburg, Mississippi, September 1999 (31 pages).

Hallum, Cecil, “Multivariate Sound Info Retrieval from MIDI Music, Digital Photography, and Satellite Imagery: SAS to the Rescue”, published in the Proceedings of the 2010 South Central SAS Educational Forum, Austin, Texas, Nov. 9-11, 2010; this is the 20th annual state-wide SAS Users’ Group Conference.

He, Huixing and Hallum, Cecil, “Multivariate Object Detection using Digital Photography and SAS-based Graphical User Interface”, *Proceedings of the Joint Statistical Meetings of the American Statistical Association* held in Salt Lake City, Utah in August 2007.

He, Nannan, *Object Detection: Search for Objects to Support Crises Aftermaths*, an M.S. practicum completed at Sam Houston State University, May 2010.

Johnson, Richard A. and Wichern, Dean W., *Applied Multivariate Statistical Analysis: Fifth Edition*, Prentice-Hall, 2002.

Minter, T.S. and Hallum, C.R., “Maximum Likelihood Classification by Thresholding”, Technical Report LEC/HASD No. 640-TR-114 under NASA Contract NAS 9-12200; appeared in *Remote Sensing of Earth Resources*, V-1, II, March 1973, pp. 795 - 808.

Minter, Jason and Hallum, Cecil, “Real-time Multivariate Analysis Onboard an RF Controlled Vehicle: Miniaturized Remote Sensing”, *Proceedings of the Joint Statistical Meetings of the American Statistical Association*, held in Seattle, Washington in August 2006.

Mitsuoka, George. “Drive a Tank BOT from your PC: Repurposed WiFi provides an elegant control solution”. *ROBOT, 1(1)*, Winter 2005, pp. 30-34.

Mohler, R.J., Hallum, Cecil, Wells, Gordon L. and Trenchard, Michael H. “Monitoring Vegetation of Drought Environments”, *Bioscience*, Vol. 36, No. 7, July/August, 1986.

Mohler, R.J., Hallum, Cecil, “Analysis of Drought Environments Using Multi-sensor Data”, *Proceedings of the Annual Meeting of the Forum of the Assoc. for Arid Lands Studies*, Reno, Nevada, April 23-26, 1986.

Perry, C.R. and Hallum, C.R., “Sampling Unit Size Considerations in Large Area Crop Inventorying Using Satellite-based Data”, *Proceedings of the Survey Research Methods Section of the American Statistical Association Conference*, August 1979, Washington, D.C.

Young, Marsha M. and Hallum, Cecil, “The Jackknifed Mahalanobis Distance Function for Change Detection in Support of Land Management Monitoring”, *Proceedings of the American Statistical Association Conference*, held in Indianapolis, Indiana, August 13-17, 2000.

APPENDIX 1: Various (Expensive) Military Drones



APPENDIX 2: Various (Less-Expensive) Drones

Various Vertical Takeoff and Landing Unmanned Aerial Vehicles



Eagle Eye



Guardian



Vigilante



Heliwing



Cypher



Wing Fan



QH-50



Dragon Warrior



Dragon Stalker



Free Wing Scorpion



Schweizer Argus



Micro Craft Lift
Augmented Ducted
Fan



Sky Technology

APPENDIX 3: Initial Miniaturized Remote Sensing Payload Carrier (for Research Purposes) – Specifics Presented at SCSUG 2007



Fig. 1: 1:6 Scale M5 Stewart Tank



Fig. 2: Linksys Wireless Router

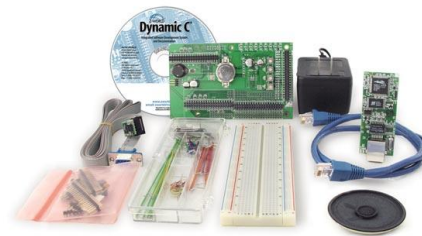


Fig. 3: Ethernet Starter Kit

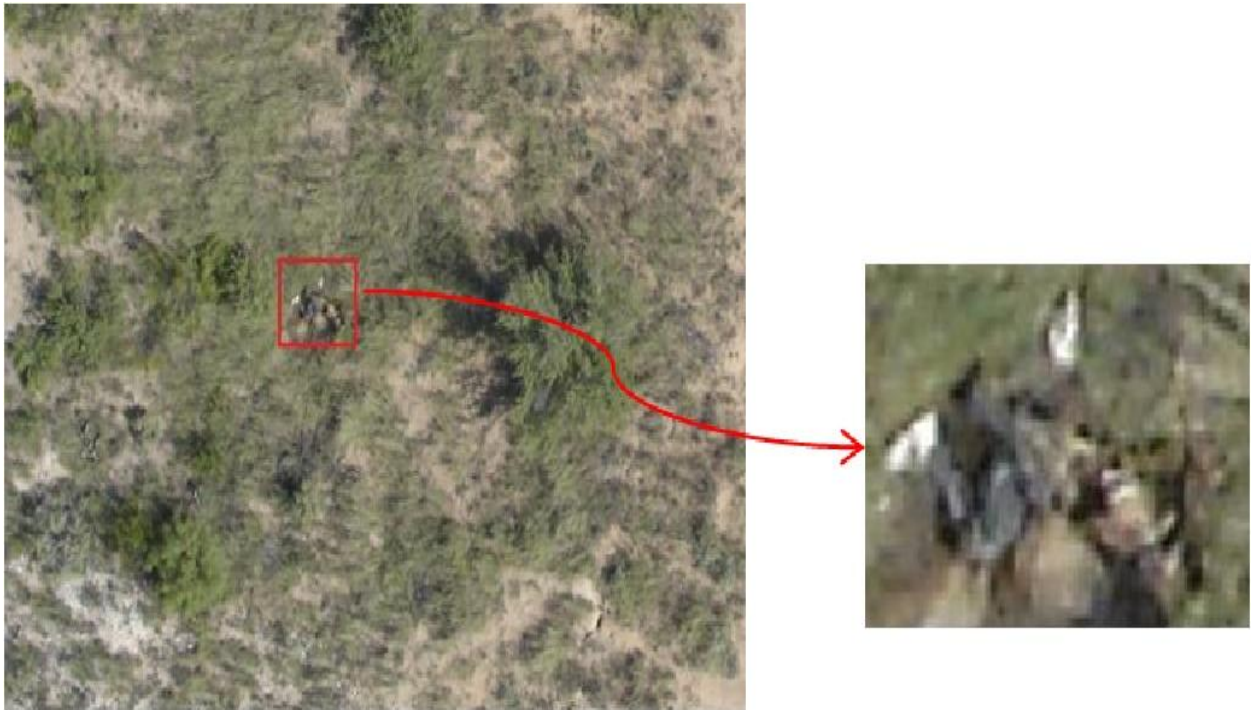


Fig. 4: Panasonic BL-C10A Webcam



Fig. 5: Replacement of Internal RC Control Board with a WiFi-controlled Platform

APPENDIX 4: Images Related to Remote Sensing Searches for Missing Bodies



Locating Objects in Digital Images

- Moving window
- Comparing to object window

