Improving Polygon Shapes Created with Landsat 7 Gap Filled Imagery

Authors: Chris Curlis, David Hansen, Barbara Simpson

Abstract

The U.S. Bureau of Reclamation employs the Central Valley Habitat Monitoring (CVHM) program to monitor changes in habitat within various areas of California's central valley. Landsat 7 ETM+ imagery is the primary data used to generate spectral based polygons and conduct an image differencing based change detection. Unfortunately imagery from Landsat 7 is affected by the Scan Line Corrector (SLC) anomaly and has systematic data gaps. However, the CVHM project uses Landsat 7 SLC-off to SLC-off Gap filled imagery. This allows us to select imagery from nearby dates (± 2 months) to fill in the gaps and create improved resolution polygons to use for change detection and classification. Areas in the imagery which still suffer data gaps are improved using imagery from adjacent scenes along with ArcGIS Model Builder and overlay tools to reduce data loss and improve the geometry of the polygons.

Project Background

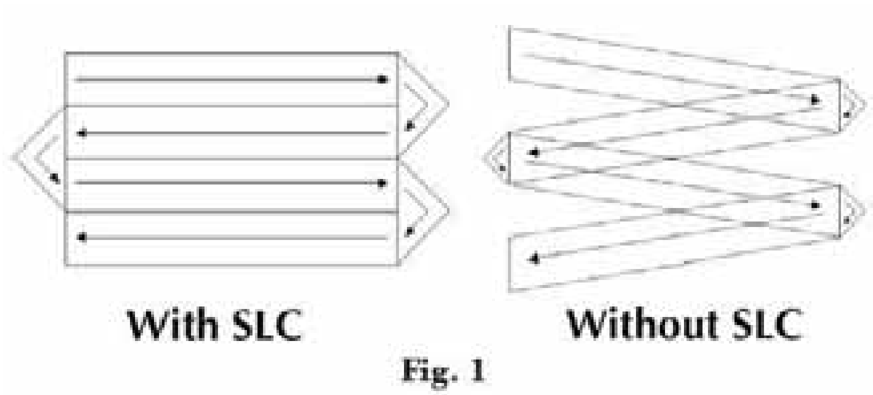
The U.S. Bureau of Reclamation (BOR) is currently conducting habitat monitoring for California's Central Valley. A program of change detection is being utilized along with the extensive review and editing of existing land cover data to identify areas of habitat loss or change.

The Central Valley Habitat Monitoring (CVHM) project utilizes image segmentation of remotely sensed imagery (Landsat TM and ETM) to create map polygons representative of land cover units on the ground. The map polygons are identified based on existing mapping data or manually edited thorough photo interpretation or field work. In the first iteration of change detection 30m Landsat 5 data from 1993 was compared to 30m Landsat 7 data for the year 2000. In this case the panchromatic imagery which is also available with Landsat 7 data was not utilized. For the next iteration we had the opportunity to use Landsat 7 data for 2005 for identifying change since 2000.

Improvements in the software available for image segmentation and the existence of Landsat 7 panchromatic 15 meter imagery (c. 2000) led us to consider using Landsat 7 pan sharpened imagery (c. 2005) for change detection and polygon generation. The improved resolution of 15 meter pan sharpened Landsat imagery over standard 30 meter Landsat imagery would allow for increased accuracy of the map polygons and thereby in the resulting maps.

Pan sharpening is an image processing technique in which a single band of high spatial resolution imagery (in this case 15 m. Landsat 7 panchromatic) is fused with multispectral imagery (the corresponding Landsat 7 multispectral data) which is resampled and transformed into modified IHS (Intensity-Hue-Saturation) bands. A further transform results in an image which combines the multispectral characteristics of the original 30m Landsat 7 data with the spatial characteristics of the panchromatic Landsat data.

Unfortunately, on May 31, 2003 the Landsat 7 satellite sensor experienced a malfunction which affected a part of the instrumentation called the Scan Line Corrector (SLC). Now known as the SLC anomaly, it has seriously affected Landsat 7 imagery by creating lost data gaps in the image. (Figure 1) All data acquired since the anomaly is referred to as SLC-off data. The existence of these gaps (Figure 2) created potential serious limitations to using the 2005 Landsat 7 data for change detection. The only other option would be to use 30 meter resolution Landsat 5 data for 2005.



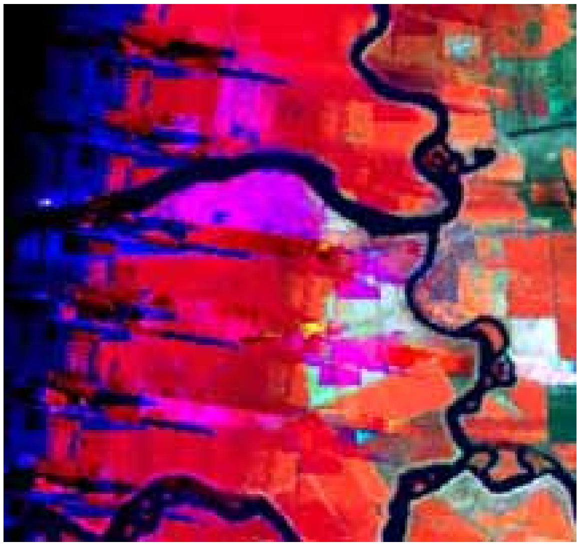
(Figure 1) Illustration of how scan line corrector (SLC) loss creates gaps in imagery
(Courtesy USGS/EROS)

(Figure 2) Example of Landsat 7 TM Imagery with gaps resulting from SLC anomaly
(Courtesy USGS/EROS)

Landsat 7 Imagery Fixes

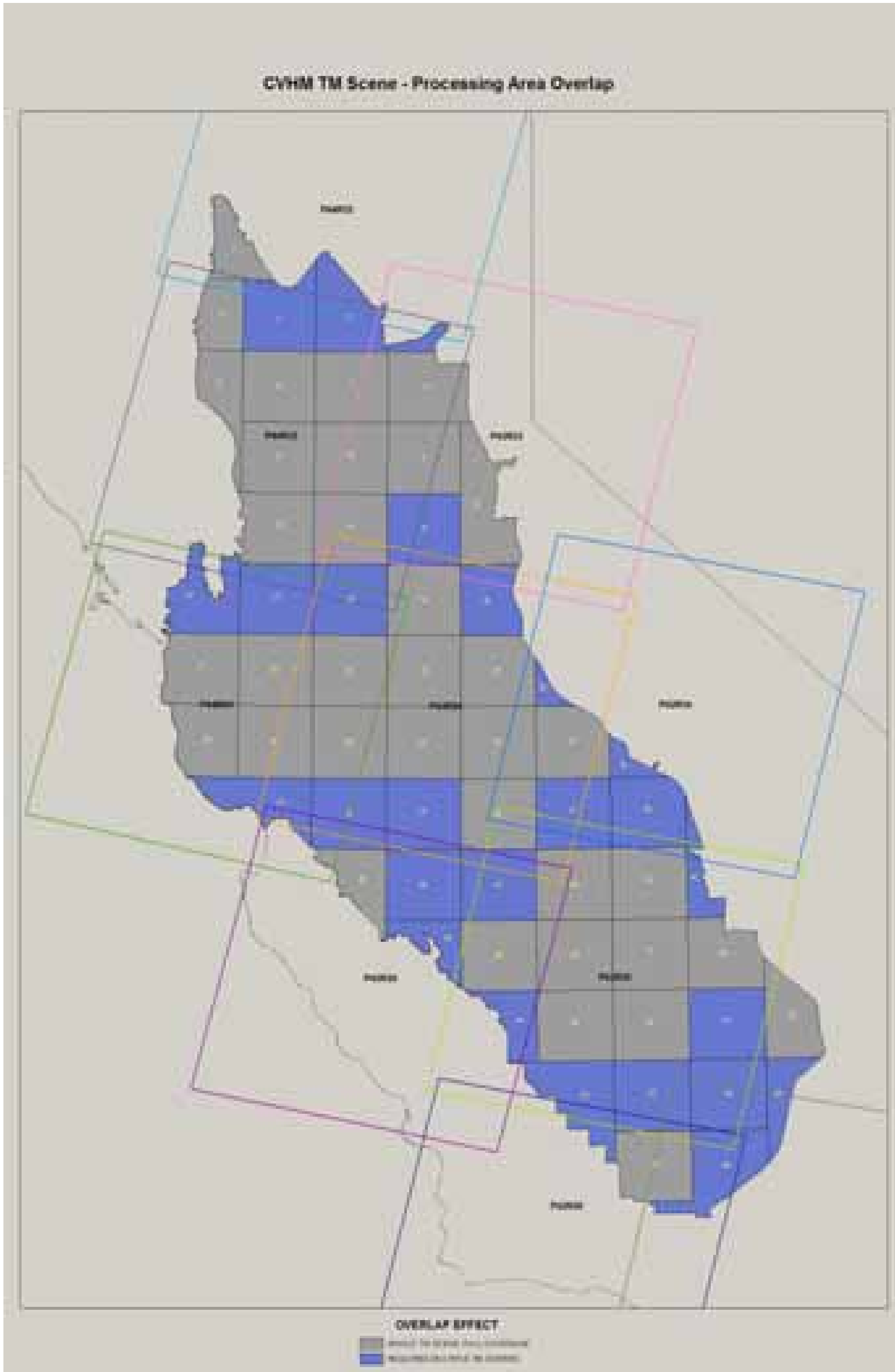
To give more utility to Landsat 7 data acquired since the SLC anomaly occurred, the U.S. Geological Survey/Earth Resource Observation System (USGS/EROS) Landsat team has devised a methodology to fill in data gaps with near-date imagery. This process uses almost exclusively SLC-off imagery and is called SLC-off to SLC-off gap filled imagery. The USGS web Global Visualization (GLOVIS) imagery ordering site (http://**glovis**.usgs.gov) allows the user to interactively select up to four alternate acquisition dates allowing for cloud free and ideally same season imagery to be used to fill the gaps in the primary image. Then, all of the missing image pixels in the original SLC-off image are replaced with histogram-matched data values derived from one or more alternate acquisition dates. Ideally these dates are close to the primary image and thus useful for broad scale change detection. For the imagery used by the CVHM project almost all alternate imagery was available near the primary imagery date (± 2 months). If the four best SLC-off dates do not completely fill the gaps then a SLC-on date (prior to May 31, 2003) is used for the remaining pixels.

The gap filling technique is most effective in the center of the Landsat scene and loses effectiveness closer to the outer side edges of a scene. At the extreme edges of a scene the gaps and actual missing or data are quite noticeable and have a dramatic effect on the polygons generated by image segmentation. (Figure 3)



(Figure 3) Example of Landsat 7 TM scene edge with extreme effects of SLC anomaly even after imagery has been gap filled.

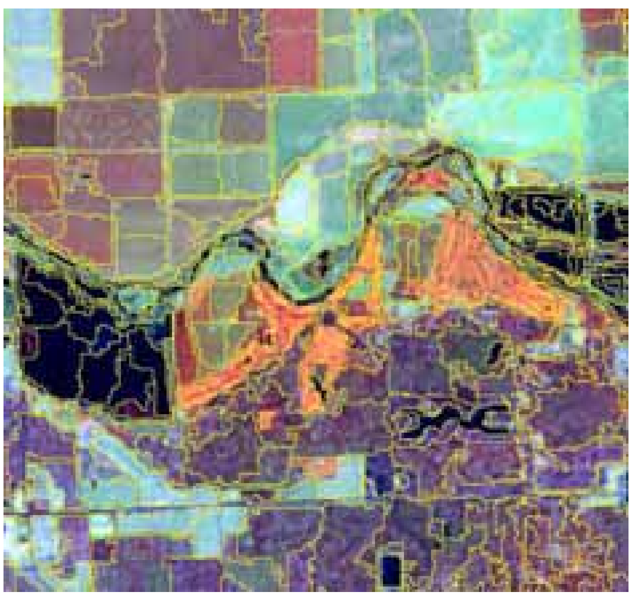
It is in these areas where a method was required to eliminate polygons formed by data gaps and then merge the remaining "good data" polygons. Landsat 7 scenes are acquired in paths as the sensor orbits the earth from North to South. The adjacent paths overlap and so there is an overlap in useable imagery along the edges of adjacent scenes. (Figure 4)



(Figure 4) Map showing Landsat 7 TM scene overlap on CVHM project area and processing areas.

Segmentation Polygons

The pan sharpened TM imagery is segmented into polygons which closely identify land use and land cover features in the project area. Definiens Ecognition is used to create polygons for an entire TM scene. A scaling factor of 25 is used in Ecognition and this generates extremely large polygon sets which require the Large Data Handling (LDH) version of the software. The typical polygons are about 25 acres in size but have a range from a fraction of an acre to a few hundred acres based on image color, texture and imagery shapes. (Figure 5) In the past we have found that a single pan sharpened TM scene is near the computational limit of our software/hardware. However, since it was more efficient for our process to create polygons for full Landsat scenes rather than subsetting the polygon creation into 58 smaller existing processing areas, we chose to create polygons by pan sharpened scene and not to use subsetted imagery or to merge multiple TM scenes prior to polygon creation due to software/hardware limitations. This required us to develop a method to merge the polygons which were created for each scene with a different set of polygons in an adjacent scene where it was necessary because of the SLC-off artifacts and to create a seamless set of polygons within each processing area.

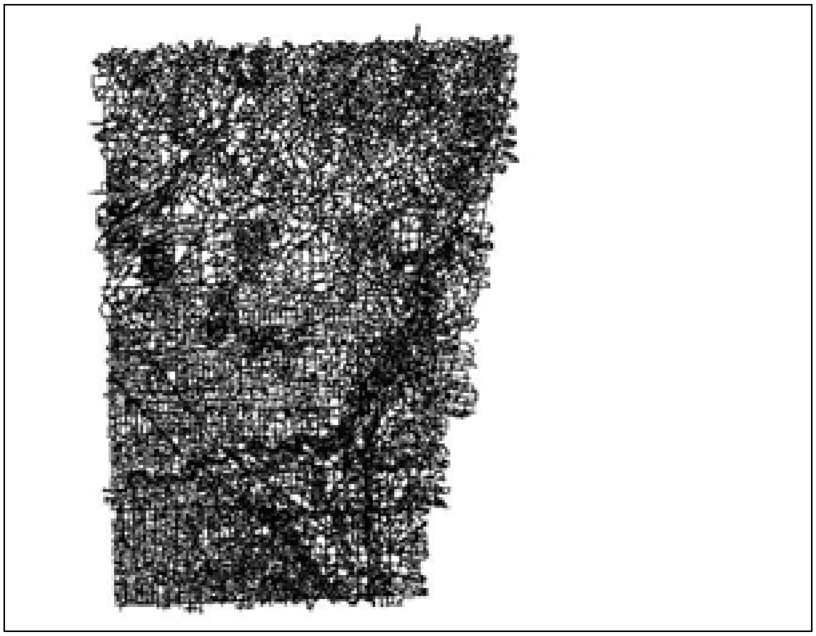


(Figure 5) A variety of polygons created based on Landsat 7 TM pan sharpened imagery.

Using ArcMap to Merge Data

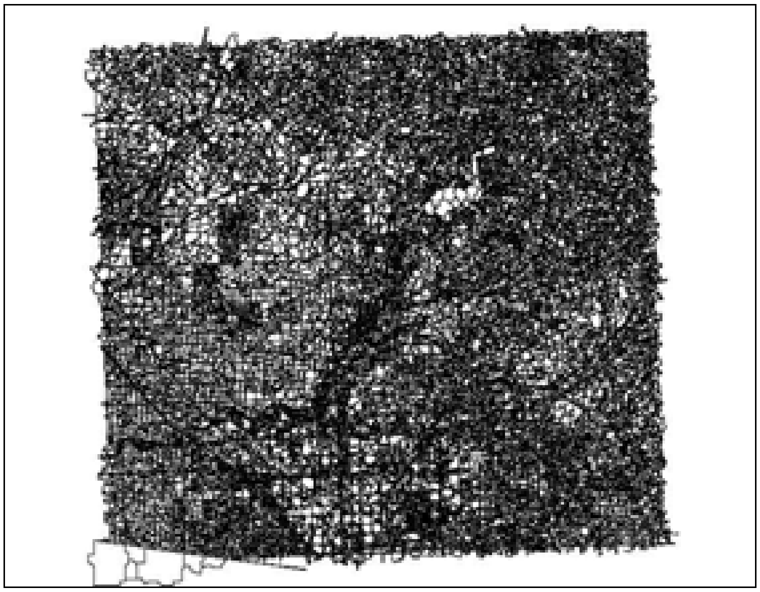
Merging adjacent polygon sets was accomplished using the **Update** tool in ArcMap in combination with an ArcMap model. The two overlapping Landsat 7 TM scenes which were used for this example are **path 43 row 34** (4334) and **path 42 row 34** (4234). The right edge of scene 4334 cut through a processing area; in this case a square processing area (**pa37**). Most of the polygons generated along this edge of imagery are misshapen by the obvious image data gaps. Scene 4234 does not fully cover **pa37** at its bottom edge. This is a perfect example of where the adjacent polygon merge is needed.

The first step involves loading the two shapefiles of polygons derived from adjacent full Landsat 7 TM scenes as well as a processing area boundary shapefile. These will all be used in the process to replace polygons formed by data gaps from the SLC anomaly. An ArcMap model was created and used to select polygons using the “select by location” option in ArcMap. It selects the polygons from TM scene path 43 row 34 (4334) which were intersected by **pa37**. The selected polygons were then exported to a feature class called **tm37_4334**. (Figure 6)

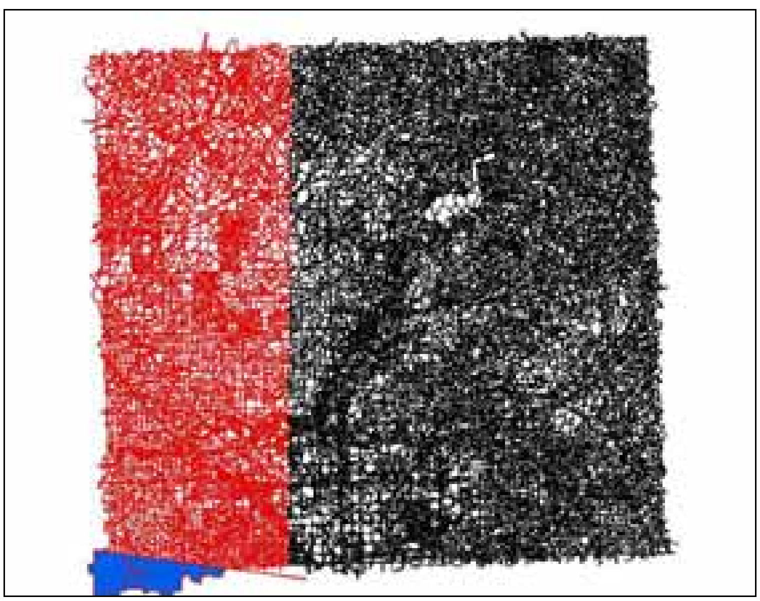


(Figure 6) The feature class **tm37_4334** showing data gaps in polygons along the right side which coincide with the edge of the Landsat TM scene.

The same process was performed for TM scene path 42 row 34 (4234) and the shapefile **pa37_4234** created. A masking field called “use” is created in this shapefile which will be used to mask out the unwanted set of polygons. (Figures 7, 7a)

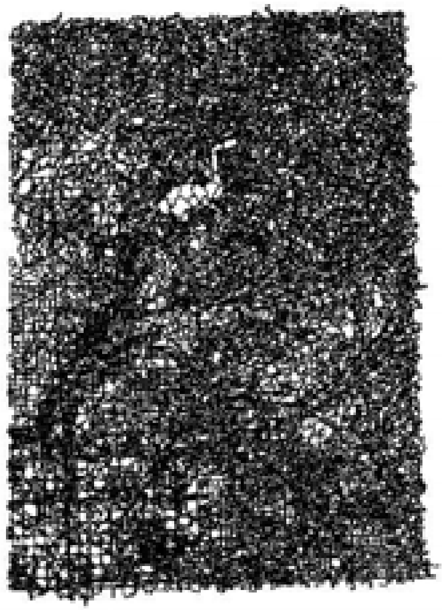


(Figure 7) The shapefile **pa37_4234**. This TM scene does not completely cover the processing area 37 as can be seen on the lower left of the polygons. This missing data was covered by polygons from TM scene 4334.



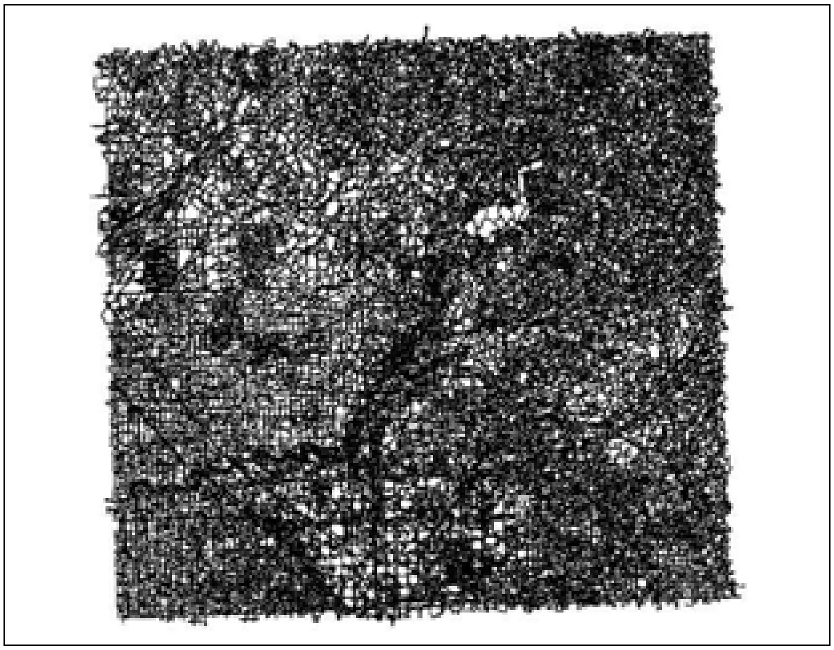
(Figure 7a) The shapefile **pa37_4234** showing masking. The red polygons indicate those which were masked and coded with "1" The blue area without red polygons on the lower left shows the area on the outside of TM scene 4234.

A vertical cut line was determined by examining imagery and polygons for scene 4234 and then all the polygons to the left of the line were selected and coded to be masked with a “1”. These polygons were excluded and the remaining polygons, coded “0” were then selected and exported to a new feature class, **tm37_4234**. (Figure 8)



(Figure 8) The feature class **tm37_4234** created by exporting the unmasked polygons from shapefile **pa37_4234**.
These polygons will be merged with the shapefile **tm37_4334** as detailed below.

The final step called for the use of the **Update** tool which is in the **Overlay** toolset of the **Analysis Tools** toolbox. This tool computed a geometric intersection of input features and update features and then used the update features to first erase and then replace the input features at the intersection.   The output was a new feature class. In this case **tm37_4334** was the input feature and **tm37_4234** was the update feature. The resulting output feature class was **tm37_42_4334**. (Figure 9) The new set of polygons has eliminated the artifacts which were created by the SLC-off imagery and thereby improve the overall quality of the mapping product.



(Figure 9) The feature class **tm37_42_4334** showing the results of the **Update** process
Between feature class **tm37_4234** and **tm37_4334**

Summary

The CVHM database for the year 2005 has been generated with pan sharpened Landsat 7 imagery which has doubled the spatial resolution available for the mapping polygons compared to earlier mapping. Ingenuity on the part of USGS/EROS allowed for the use of Landsat 7 TM data for change detection in spite of its shortcomings. By using improved image segmentation software in combination with innovative methods for polygon merging and data fusion, the 2005 update as well as future iterations of habitat mapping will move closer to an approximation of reality.

References

SLC-off Products: Background, U.S. Geological Survey, Washington, DC. 2005, http://landsat.usgs.gov/data_products/slc_off_data_products/slc_off_background.php.

Chris Curlis
David T. Hansen
Barbara Simpson
MPGIS
U.S. Bureau of Reclamation
mid Pacific Region
2800 Cottage Way

Sacramento, CA. USA 95825-1898
Phone: (916) 978-5030
FAX: (916) 978-5290
Email: ccurlis@mp.usbr.gov