Local Search for Optimal Global Map Generation Using Mid-Decadal Landsat Images

Robert A. Morris[2] and **John Gasch**[3] and **Lina Khatib**[1,2]
[1] SGT
[2] Computational Sciences Division, NASA Ames Research Center
[3] Landsat Mission Operations, Goddard Space Flight Center

Abstract

NASA and USGS are collaborating to produce a global map of Earth using Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper Plus sensor data from the period of 2004 through 2007. The map is comprised of thousands of scene locations and, for each location, there are tens of different images of varying quality to chose from. Constraints and preferences on map quality make it desirable to develop an automated solution to the map generation problem. This paper formulates a Global Map Generator problem as a *Constraint Optimization Problem* (GMG-COP) and describes an approach to solving it using local search. The paper also describes the integration of a GMG solver into a user interface for visualizing and comparing solutions.

Introduction and Motivation

NASA's LCLUC[1] Program is partnering with the EROS[2] Data Center to produce a high resolution mosaic map of the Earth. The map will consist of a data set of high quality images of the Earth's continental landmass using Landsat 5 (L5) Thematic Mapper (TM) and Landsat 7 (L7) Enhanced Thematic Mapper Plus (ETM+) sensor data from the mid-decadal period of 2004 through 2007. This project is known as the *Global Land Survey*, or GLS-2005.

The end-product will be composed of roughly 9500 Worldwide Reference System 2 (WRS-2) [3] Landsat scene locations, of which there are typically ten or more high quality candidate images available for each scene location. Eventually, over 300,000 images must be evaluated and down-selected to create the final survey data set. The resulting data map will be distributed to the public at no charge through a USGS website. In addition to providing benefits to researchers in the Earth sciences, it will likely become the next generation backdrop for Google-Earth (which currently uses the GeoCover-2000 data set).

A collection of diverse preference criteria defines a high quality image map. First, a good map will typically consist of the best (most cloud-free) image data available per scene. The metric employed for this measure is the Automated Cloud Cover Assessment (ACCA), a statistic derived from an algorithm that identifies clouds from data through difference in mean temperature with Earth surface. Second, as the majority of Earth Science applications deal with health and density of agricultural and other vegetative land cover, images taken during peak vegetation maturity are preferred. The metric employed for this purpose is the Normalized Difference Vegetation Index (NDVI). NDVI represents the historical average vegetation density and maturity within each global land scene cell by calendar month. Thus, the preference is toward images taken during seasons with high NDVI. Third, to be usable for regional scientific studies, it is preferable to choose image data that are seasonally consistent with neighboring scenes. Fourth, to accommodate land-cover/land-use change analysis, consideration must be given to the seasonality of previous survey data sets, so images taken in the same season are preferred. Finally, due to a malfunction of the image scanner on L7, ETM+ produces imagery that has coverage discontinuities such that an individual image covers only 78% of the land area. To compensate, two images of the same scene taken on different days must be combined to produce a composite image that partially or fully closes the gaps. Pairs of images of a common scene must therefore be chosen to maximize coverage (minimize gap), which means the two scenes should be mutually out of phase. Each image is assigned a "gap–phase statistic", or GPS, which is an absolute measure of the geometric registration of the image scan line with respect to the scene center point. Such GPS values are used to compute the area coverage of composite images.

The size of the space of global image maps, as well as the number of criteria for quality, make it desirable to automate the production of solutions to the problem. This paper presents a formulation of the map generation problem as a Constraint Optimization Problem, and describes an approach to solving the problem using local search. Section 2 describes the problem in more technical detail. Section 3 describes a local search approach to solving the problem.

[1] Land-Cover and Land-Use Change http://lcluc.umd.edu.

[2] USGS Earth Resources and Observation Science http://eros.usgs.gov.

[3] L5 and L7 follow the WRS-2 coordinate system for indexing locations on the Earth where data is acquired. WRS-2 indexes a location via a set of paths and rows, with a 16-day repeat cycle. L5 follows the WRS-2 system with a temporal offset of 8 days relative to L7. The WRS-2 indexes orbits (paths) and scene centers (rows) into a global grid system (day time and night time) of 233 paths by 248 rows. We refer to each path, row element as a *scene location*.

Section 4 discusses the user interface, Large Area Scene Selection Interface (LASSI) into which the solver is integrated.

Global Map Generation as a Constraint Optimization Problem

Global Map Generation (GMG) can be viewed as a *Constraint Optimization Problem* (GMG-COP) (Larrosa & Dechter 2003), with a set of variables $V = \{v_{i,j}\}$ indexed by WRS-2 path and row number i, j. Each variable $v_{i,j}$ represents a scene location, and is associated with a *domain* $D_{i,j} = \{d_{i,j,1}, \ldots, d_{i,j,m}\}$, where each $d_{i,j,k}$ represents a TM or ETM+ image taken of the corresponding scene. The WRS organization of scene locations into path and row induces a grid or lattice structure of binary relations between scene locations. Specifically, there are binary links with 4 neighboring scenes, designated $north(v_{i,j-1})$, $east(v_{i-1,j})$, $south(v_{i,j+1})$, and $west(v_{i+1,j})$.

A *solution* s to the GMG-COP is a set of assignments $s = \{v_{i,j} \leftarrow \langle d_{i,j,k}, d_{i,j,l} \rangle\}$. The need for a pair of images arises from the L7 ETM+ gap anomaly. One partial image is called the *base*, which covers approximately 78% of the WRS scene area. The other partial image is called the *fill*. If the base is a TM image from L5, where there are no missing image data, we set $d_{i,j,l} = d_{i,j,k}$ by convention. For an arbitrary solution s, we write $b_s(v_{i,j}), f_s(v_{i,j})$ for the base and fill values for the scene $v_{i,j}$ assigned by s.

The GMG problem is a multiobjective optimization problem, in which a set of potentially competing preference criteria are used to evaluate and compare solutions. The preference criteria are the following:

1. Single image criteria:
   - Minimize cloud cover.
   - Maximize NDVI value.
   - Seasonality with previous data sets.
   - Relative preference for acquiring L7 versus L5 images.
   - Reward for acquisition dates centered in study period (2005 or 2006 versus the fringe years 2004 or 2007).
2. ETM+ composite criteria:
   - Minimize gaps in data that remain after compositing image pairs.
   - Minimize the temporal difference between the composited images acquisition dates.
3. Criteria relating pairs of adjacent images:
   - Minimize the temporal difference between the image acquisition dates.
   - Minimize the seasonality difference between the images (the days of the year, ignoring the year, the images were acquired).
   - Prefer adjacent images acquired from a common sensor (sensor homogeneity) i.e. both TM or both ETM+.

For the purpose of generating a global image map, each candidate image is represented as a vector of "meta-data" comprised of the following text fields: the WRS-2 scene path and row numbers, the sensor that acquired the image (TM or ETM+), the acquisition date, the cloud cover assessment ACCA, the NDVI metric, and a GPS value used for evaluating the scene area coverage that results from compositing two images. In addition, a preference date is specified for each scene location, which corresponds to the acquisition date of the same scene within the earlier GeoCover-2000 data set. This preference encourages solutions for GLS-2005 that are temporally compatible with GeoCover-2000, thus enabling comparitive change research between the two data sets.

Each meta-data vector element is associated with a function that is used to evaluate solution quality. We normalize by considering *merit* values in the range $[0, 1]$ where 0 is worst and 1 is best. This way the objective function is a maximization and always positive. First, the quality of an individual image can be depicted in terms of two functions: NDVI merit value: $ndvi : D \rightarrow [0, 1]$; and cloud cover: $acca : D \rightarrow [0, 1]$. Second, there are two functions associated with measuring the time difference between the acquisition of neighboring pairs of images. Absolute Day Difference, $date\Delta : D \times D \rightarrow [0, 1]$ is the number of days between image acquisition. Day of Year Difference, $doy\Delta : D \times D \rightarrow [0, 1]$ is the gap in days (ignoring the year in which it was acquired). The latter function is used to reward solutions that assign images that are seasonally similar, regardless of year, whereas the former rewards solutions with pairs of images with small temporal distance. Third, the function *Area Coverage*, $cover : D \times D \rightarrow [0, 1]$ assigns a value that indicates goodness of fit between a base and fill image used in a composite. Finally, to express relative preferences for TM or ETM+ images, the function $IsL5 : D \rightarrow \{0, 1\}$, $IsL7 : D \rightarrow \{0, 1\}$ assign 1 to images acquired by TM (respectively, ETM+), and 0 otherwise.

Because of the symmetry of adjacency, it suffices to represent the set of adjacent scenes in terms of the functions $north : V_{i,j} \rightarrow V_{i,j-1}$ and $east : V_{i,j} \rightarrow V_{i-1,j}$, which return the variable corresponding to the scene that is north (east) of the designated variable.

The set of solutions can be ordered in terms of the objective of maximizing individual scene quality while maximizing phase difference between bases and fills, and minimizing the temporal differences between (the bases of) adjacent images. Given an arbitrary solution s, its score is the value of the following weighted summation:

$$
\begin{aligned}
f(s) \quad = \quad & \Sigma_{i,j} \\
& w_1 * ndvi(b_s(v_{i,j})) \\
+ \quad & w_2 * acca(b_s(v_{i,j}))) \\
+ \quad & w_3 * ndvi(f_s(v_{i,j})) \\
+ \quad & w_4 * acca(f_s(v_{i,j})) \\
+ \quad & w_5 * date\Delta(b_s(v_{i,j}), f_s(v_{i,j})) \\
+ \quad & w_6 * cover(b_s(v_{i,j}), f_s(v_{i,j})) \\
+ \quad & w_7 * date\Delta(b_s(v_{i,j}), b_s(north(v_{i,j}))) \\
+ \quad & w_8 * date\Delta(b_s(v_{i,j}), b_s(east(v_{i,j}))) \\
+ \quad & w_9 * doy\Delta(b_s(v_{i,j}), b_s(north(v_{i,j}))) \\
+ \quad & w_{10} * doy\Delta(b_s(v_{i,j}), b_s(east(v_{i,j}))) \\
+ \quad & w_{11} * IsL5(b_s(v_{i,j}) \\
+ \quad & w_{12} * IsL7(b_s(v_{i,j})
\end{aligned}
$$

w_1 and w_2 govern the importance of the quality of individual base images. $w_3 - w_6$ discount the value of an image based on the quality of the fill, and on the goodness of fit between base and fill. For L5 images, where the base and fill are the same, the discount is the same as $(w_2 + w_4)acca(b_s(v_{i,j}))$, etc. Notice that since we assume that the temporal and spatial match between an image and itself is perfect, L5 images are not discounted on these criteria. $w_7 - w_{10}$ deal with compatibility of bases with adjacent images (we ignore the compatibilities of fill), and w_{11} and w_{12} allow for an absolute preference for L5 or L7 images to be expressed. An optimal solution s^* to this GMG problem is one that receives the maximum score based on this function.

Solution Using Local Search

Local search defines a class of approximation algorithms that can find near-optimal solutions within reasonable running times. Given the pair (S, f), where S is the set of solutions and f is the objective function, let $S*$ be the set of best solutions (i.e., the ones with the highest score according to f), and $f*$ be the best score. Members of $S*$ are called *global optima*. Local search iteratively searches through the set S to find a *local optimal* solution, a solution for which no better can be found. Local optima need not be in general members of $S*$.

Complete Approaches to Solving COPs

Computationally, the problem solved by GMG is similar in structure to the problem of assigning frequencies to radio transmitters (Cabon *et al.* 1999), and other generalizations of the map coloring problem. There are two complete general constraint-based methods for solving such problems: through search, as with Branch and Bound algorithms; and through variable elimination, e.g. using Bucket Elimination (Dechter 2003). The worst case time and space complexity of the latter is tightly bounded by a parameter of the problem called the *induced width*, which arises out of an ordering of the variables. Specifically, complexity of Bucket Elimination is $O(n * d^{w+1})$, where n is the number of variables, d is the size of the largest domain, and w is the induced width. In practice, the primary drawback in performance is space; only problems with small induced width can be solved.

Given a set of variables and associated constraints, finding the ordering of the variables with a *minimum induced width* is an NP-hard problem. Although to our knowledge no proof exists, it appears that the induced width of a constraint graph arranged as a square grid of size $n \times n$ is n. This linear growth rate imposes a practical limitation on the size of problems solved in a reasonable time by Bucket Elimination to roughly $n = 30$, too small for the GMG problem (Larrosa 2007). Hybrid approaches that combine Bucket Elimination with Branch and Bound have demonstrated an improvement in performance over pure Bucket Elimination for problems with a grid structure (Larrosa, Morancho, & Niso 2005). Although these results justify the future application of these methods to the GMG problem, in this effort we did not attempt to solve the problem using a complete method, but rather chose local search.

Reasons for Local Search

Reasons for adopting a local search method to solving constraint optimization problems are well documented: they include

1. *Anytime performance:* On average, local search behaves well in practice, yielding low-order polynomial running times (Aarts & Lenstra 1997). Because the criteria space is high-dimensional, it is difficult *a priori* to quantitatively characterize globally preferred solutions. Consequently, our customers were interested in a system that could examine large parts of the search space quickly to determine weight settings that produced adequate results.
2. *Flexibility and ease of implementation:* Our customers required us to build, and demonstrate the advantages of, automated solutions in a short period of time (2 months). Local search can be easily implemented.
3. *Ability to solve large problems:* As optimization problems go, the GMG-COP can be considered large. Local search has been shown to be effective on large problems.

There are also domain-specific reasons for choosing local search. Specifically, since the cloud assessment (ACCA) statistic is a close but imperfect metric of "cloud truth", particularly around coastlines, it was felt that it made more sense to have a solver that could generate multiple solutions easily and allow humans to conduct further evaluations of the solutions and, where necessary, "tweak" them. A complete solver, which would likely take significantly longer in generating a single solution, would make such an iterative process less viable.

Implementation

To conduct the search, local search relies on the notion of a *neighborhood function*, $N : S \rightarrow S$, a function that takes one solution (called the *current solution*) and returns a new solution that differs from the current in some small way. To be effective, a neighborhood function should be simple; it should not require a lot of time to compute. For GMG-COP, the neighborhood function randomly selects a cell and replaces the selected image with a new one. A neighborhood function is *exact* if every local optima it finds as the result of search is a global optima. The neighborhood function used to solve GMG-COP is not exact.

Designing a local search algorithm is based on deciding three components: how an initial solution, or *seed*, is generated, how to select a neighboring solution, and when to terminate search. For the GMG solver described in this paper, we took a simple approach to deciding these issues, reasoning that complexity should be introduced only as needed, i.e., only as warranted by inferior performance of simpler approaches, as expressed by the customers.

First, a good design for a seed generator is one that intuitively *starts in a good location* in the search space. A good location is one that is relatively close to optimal solutions, where close is measured by the length of the path from it to an optimal solution using the neighborhood function. For the GMG-COP, we chose a seed that picks the highest individual quality image for each cell, ignoring preferences related to adjacency. This seed is easy to generate (there is no need to consider adjacency constraints) and should be a good quality solution because it favors cloud-free images with high NDVI value.

Second, choosing a neighboring solution requires, first, choosing which cell to change. The simplest approach is to pick the cell at random. Since local search is "memoryless", in the sense that it does not keep track of where it's been previously, it may not be able in general to avoid examining the same solution multiple times. To avoid this, sometimes algorithms have "taboo" lists, lists of variables recently chosen to change. Variables are put on the list after being chosen and eventually taken off after some number of iterations. Variables on the list can't be selected on a given iteration. In our implementation we applied an extreme case of "taboo" list: once a scene is selected for examination, it is immediately placed on the taboo list to allow for all other scenes to be examined in the current iteration (the ordering of scene selection is random).

Third, given a selected cell, there are also a number of ways to select among the set of neighboring solutions based on changes made to that cell. Some are deterministic; i.e., given the same decision to make, the algorithm will make the same choice each time. Others are non-deterministic. Algorithms such as simulated annealing and genetic algorithms are non-deterministic. Initially, we opted for a deterministic approach, of which there are two kinds: first improvement or best improvement. First improvement examines neighbors, in a local search sense, until one is found that is better than the current solution; that one becomes the new current solution. Best improvement examines all the neighbors, and picks the one that improves upon the current solution the most. Either of these generates a *greedy* approach, one that always chooses an improving solution. A variation of best improvement is where a neighbor with the best score is chosen, even if the score is worse than the score of the current solution. This approach allows for the possibility that a globally optimal solution may not be on the "greedy path" from an initial seed solution.

Finally, choosing a termination condition requires deciding how many solutions will be generated before the algorithm halts. The simplest approach will be to define a termination condition that says *halt when you reach the first locally optimal solution or after a fixed number of solutions, MAX, have been generated*, whichever comes first. A slightly more sophisticated version of this *simple local search* is called *multi-start*: here, for some fixed number of runs, we start with different initial (seed) solutions. Such initial solutions can be fully randomly generated (our implementation), semi-randomly generated, or deterministically generated. An example of deterministically generated initial solutions employed here is to assign to each scene the best self-quality image/pair. Alternatively, the local optimum of one run of simple local search can be used as the initial solution for the next run.

Testing and Results

Testing the GMG-COP occurred in two stages. First, we compared different variations in multi-start local search to determine the best performing algorithm. Four variations were tested, based on two variations of two criteria: the initial solution and the choice of neighbor. The initial solutions tried were a randomly generated solution and the solution consisting of the set of images that scored highest individually (i.e. with respect to cloud cover, NDVI, and base–fill quality). The choice of neighbor was either done on a "first improvement" basis, i.e., the first alternative that improved the overall score, or "best improvement" basis, i.e., of all the images, selecting the one that most improved the score. The results indicate that the best strategy for finding high quality solutions is through exploration: with a random initial solution, and a "best improvement" neighbor selection, progress was quickly made towards solutions with higher quality than those found by the other approaches. We speculate that a random seed works better than one based on individual scene quality because the latter forced the search into local optimum that was not globally optimal.

In the other stage, we were interested in the extent of the improvement offered by an automated solution over current practice, which consists of manually generating solutions. Towards this end, tests were conducted by the customers at USGS and the Landsat mission using the GeoCover-2000 (GC2K) data set. The results showed that GMG, implemented as a simple algorithm then which we later improved significantly, produced a solution that was 23% better quality than the manually generated solution, based on the objective function scores. The customers viewed this result as significant enough to warrant integration and deployment of the solver.

On the complete GLS-2005 data set, the GMG solver converges to a solution in about a minute. A more detailed discussion of experiments during GMG development is found in (Khatib *et al.* 2007) (Morris *et al.* 2008).

Large Area Scene selection Interface (LASSI)

As noted above, the GMG solver arrives at its solution based on input consisting of a metadata representation of an image. Metadata furnish a low-fidelity, quantified assessment of several image criteria, such as cloud contamination and vegetation maturity (NDVI). Each metadata metric is a global average assessment over the entire image area, but each metric is subject to its own systematic errors. For example (Franks *et al.* 2008), the ACCA algorithm for generating cloud-cover metadata sometimes is not able to detect prevalent cirrus clouds, haze, or forest fire smoke. Visual inspection, on the other hand, can easily detect these scene imperfections. In general, a GMG solution is only as good as the metadata it uses to select the corresponding image, so noisy input can translate into a less than ideal solution. To address the reality of these potential sources of sub-optimal solutions, GMG is embedded into a graphical user interface and visualization tool known as "LASSI" (Large Area Scene selection Interface)[4]. LASSI allows users to:

- Adjust objective function criteria weights prior to launching the GMG solver.
- Visually assess mosaic thumbnail image renderings of the GMG solutions.
- Examine quality of solutions with respect to each objective function criterion.
- Manually disqualify individual images from consideration by GMG based on visual assessment.
- Manually override images selected by GMG based on visual assessment and external knowledge.
- Remotely access and view browse imagery for each WRS cell from the USGS Landsat image archive.

In the following, we illustrate these capabilities in the process of scene selection for GLS2005, summarizing some of the details found in (Franks *et al.* 2008).

Setting Weights for GLS2005

Scene selection for GLS2005 often involved running GMG separately for different regions of the Earth, with different weight assignments to criteria based on different climate and other relevant differences among the regions. The user initializes these weights somewhat by educated trial and error. The relative weights are influenced by the depth of the candidate image pool, the land cover characteristics, and seasonality. A deeper candidate pool allows the user to set the weights more aggressively. In dry regions, the ACCA weights can be relaxed in favor of more aggressive NDVI. Conversely, in tropical regions the NDVI and temporal factors can be downweighted in favor of more aggressive ACCA. For agricultural regions, we prefer L5 TM where available. For other temperate regions, the temporal factors are more important, trading slight degradation of ACCA and/or NDVI in favor of selections that minimize bordering scene discontinuities.

For example, Table 1 shows weight assignments eventually chosen for North American scene selection. This assignment reflects the relative importance of the vegetation (NDVI) criteria, due to the primary application of the data for land-use change studies. Cloud cover weights (ACCA) could be relatively lower because the average cloud cover of the data set was low, and hence this criterion could be fairly easily satisfied. The two distinct values for acquisition date difference between base and filler images arose from the importance of a complete image in areas of rapid land-use change, such as croplands. In general, the goal for GLS2005 data was to achieve a 95% coverage with each base-pair. For more details on the interpretation of the weight assignments, see (Franks *et al.* 2008) which focuses in depth on the user perspective.

[4]LASSI is currently tailored toward Landsat imagery, but it is feasible to customize it to work with other imagery sources.

Viewing Solution Quality Maps

The LASSI interface allows users to examine the quality of each GMG-generated solution on each individual criterion. Each *meta-data map* portrays a metadata criterion in color gradients on a WRS-2 map grid. These maps enable the user to assess the quality of the solution with respect to a single dimension. Figure 1 is one such view, in this instance, of the NDVI metric. Similar metadata maps are available for

- Sensor - Discriminates Landsat 5 TM versus Landsat 7 ETM+ images in the solution set.
- Day of Year - Relative time of year of base acquisitions.
- Year - Acquisition year.
- ACCA - Cloud assessment, measure of cloud pixels (domain 0 to 10 %).
- NDVI - NDVI, normalized with respect to peak NDVI by WRS scene.
- Raw NDVI - Non-normalized (raw) NDVI.
- Preference Year - Distinguishes whether the chosen image was acquired within the middle two years (2005 or 2006) or the "fringe years" (2004 or 2007).
- Preferred Day of Year - Depicts the seasonal temporal difference between acquisition dates of the chosen image with respect to the date of the corresponding WRS scene in the GeoCover-2000 data set. (Minimizing this time difference improves the utility of the two data sets for trending analysis.)
- Northern Neighbor - Depicts the temporal difference between neighboring "along-track" scenes (north to south). Minimizing this difference reduces potential discontinuities in the resulting end-product map.
- Eastern Neighbor - Depicts the temporal difference between neighboring "adjacent-path" scenes (east to west).

Other matadata views depict the quality of images chosen to fill gaps in the L7 ETM+ images. These include quality of the ACCA fill, the cloud assessment of gap fill images; coverage, i.e., the relative success of gap-filling (100% coverage is desired for each base-fill image pair); NDVI difference, i.e., for each base-fill image pair, the relative difference in seasonality between these images (If the difference is too great, the pair may produce a composited image with undesirable "artifacts"); and temporal difference, i.e., for each base-fill image pair, the relative difference in acquisition dates between these images (if the difference is too great, the pair may result in artifacts similar to NDVI differences, in addition to differences in sun angle).

Weight Value	Criteria Description
60	NDVI Base Image
30	NDVI Fill Image
20	ACCA Base Image
20	ACCA Fill Image
10	Gap-filler acquisition date difference
40	Gap-filler difference (over agricultural areas)
15	Area coverage
4	Day of year difference between north/south neighbors
4	Day of year difference between east/west neighbors
0	Preference for L5 imagery
10	Preference for L7 imagery
5	Sensor homogeneity
10	Preference for specific date range (2005 and 2006)
15	Preference for day of year (with respect to Geocover2000 data set)

Table 1: Relative weights of criteria used for generating scenes for North America using GLS2005 data set (Franks *et al.* 2008).

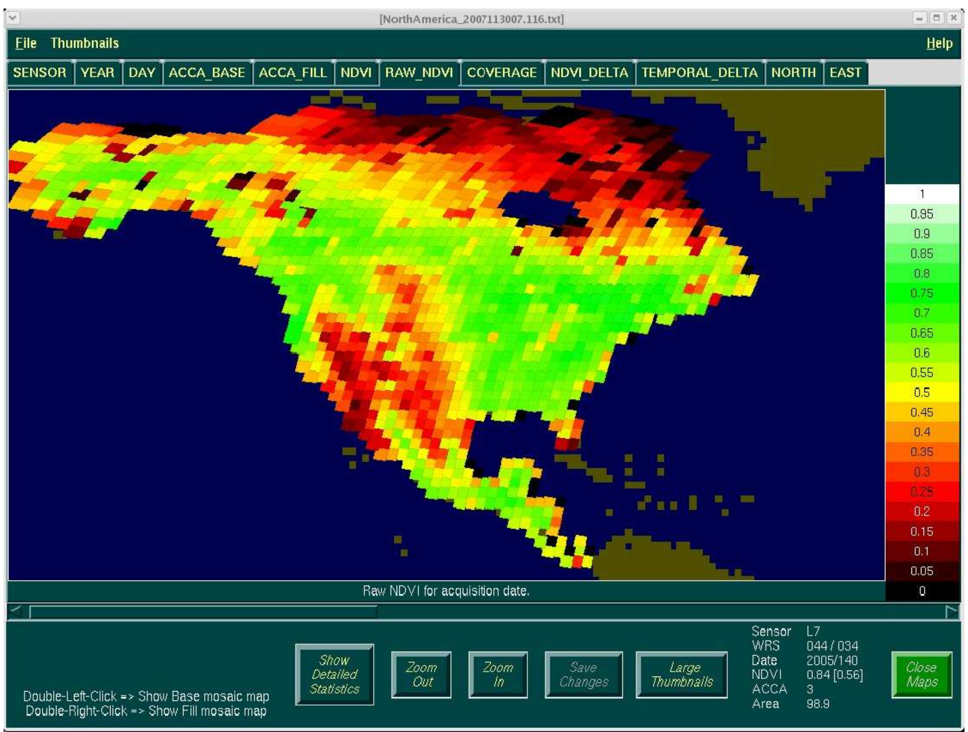

Figure 1: LASSI screen capture showing NDVI quality of the GMG-generated solution after optimizing over all criteria. Lighter color signifies better quality image of the displayed criterion.

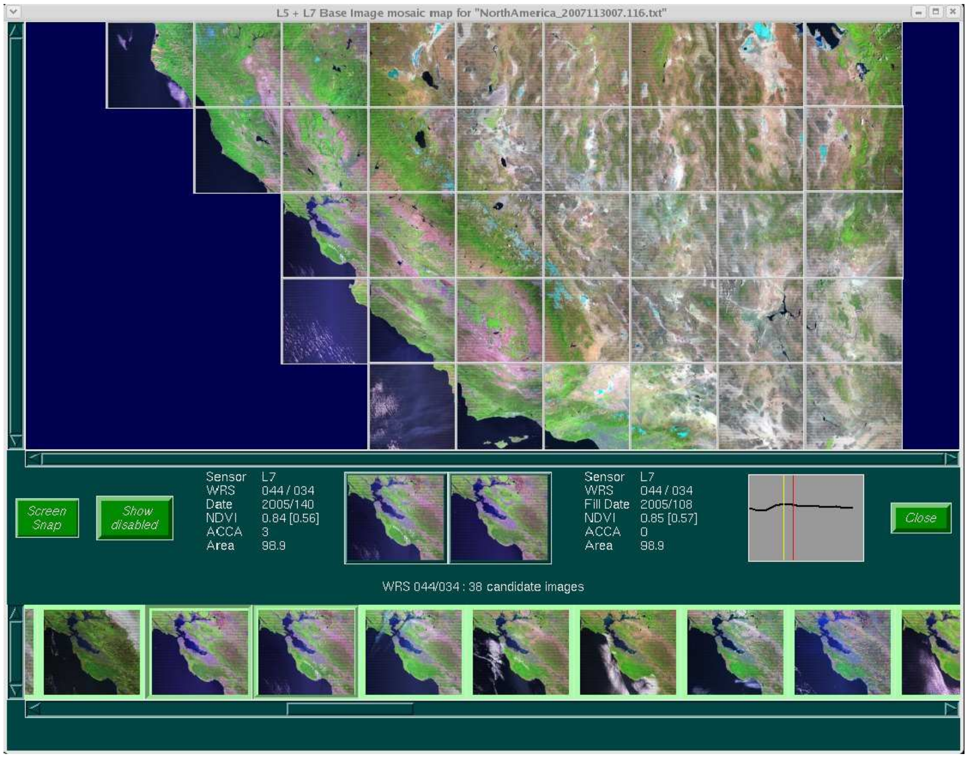

Figure 2: Screen shot of thumbnail images of Southwest United States; a portion of GMG-optimized solution of North America. For a selected thumbnail, the LASSI interface shows images not selected at the bottom of screen. User may override GMG selected image by choosing one of the alternatives.

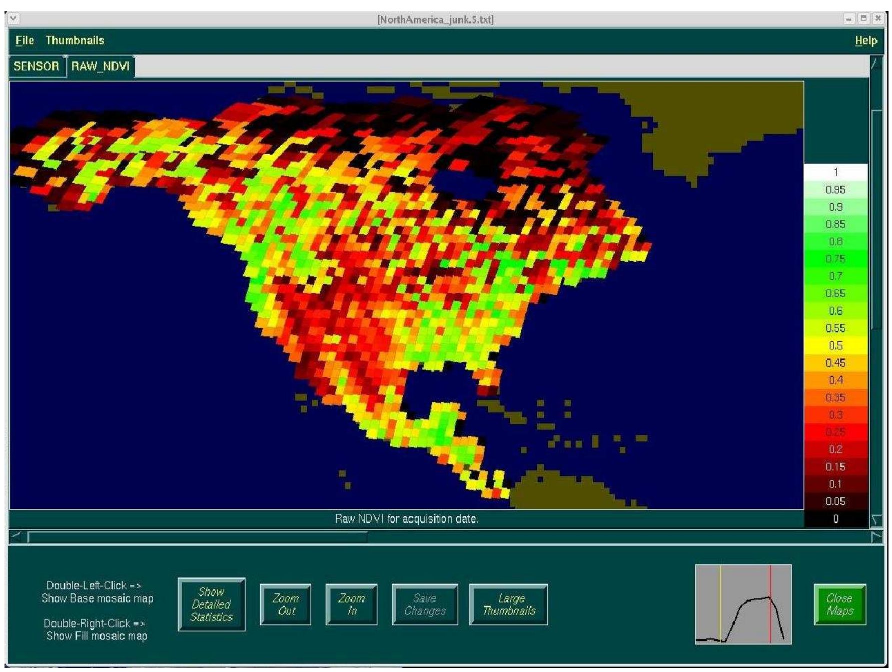

Figure 3: LASSI screen capture showing North America NDVI quality degradation when GMG optimizes only Cloud Cover (ACCA) criterion without regard for NDVI or temporal criteria.

Viewing Thumbnail Image Mosaics

After a solution has been produced by GMG, the user may use the LASSI interface to view thumbnails of the selected scenes. By double-clicking any metadata map, LASSI produces a mosaic map of thumbnail browse images for that area. The user may also view a full-screen browse image of any thumbnail. This is especially useful for revealing cloud contaminated images that may not be accurately represented in the metadata ACCA criterion, as explained above.

This mosaic map display also enables the user to view the chosen ETM+ base and fill images for side by side comparison, as seen in Figure 2. A small window in the lower right of this display plots the monthly NDVI of this scene with markers showing the relative acquisition time of year of the base and fill images (where applicable). Adjacent to each thumbnail is a list of metadata criteria and values. Finally, the bottom of the screen features a horizontally scrolling list of thumbnails for all candidate images of any selected WRS grid cell. From this list, the user may manually override the original GMG selection by choosing alternate acquisitions for the base and/or the fill images.

Global Scene Generation Process

Global scene generation using GMG/LASSI is an iterative process. The user initializes the objective function weight parameters and then invokes GMG to produce a strawman solution. After examining the metadata maps, the user tweaks these weights if necessary to compromise in some dimensions to improve others. In the end, the user will view the solution's thumbnail mosaic map and manually fine-tune it if necessary to eliminate imagery with popcorn clouds, contrails, snow, or other contaminents that may not have been accounted for in the metadata.

For the 2005 global map construction, the project has elected to pursue the problem independently for each continental landmass. This way the GMG objective function weights may be tailored for each global region. The 2005 scene selection of North America using GMG/LASSI was completed first. Then, by the end of 2008, the scene selection of the rest of the global continental landmass was completed. Availability of data to date can be found at http://landsat.usgs.gov/science_GLS2005.php.

With multi-criteria optimization, we often encounter the problem of the negative interactions among criteria, i.e., when increasing the quality of a solution with respect to one criterion causes a decreased quality with respect to another. For example, when only the ACCA criterion was preferred, GMG generated a map for North America that is suboptimal in NDVI. See Figure 3 and compare it to Figure 1 where optimization occurred jointly over both factors. Note the existence of darker areas which indicates lower quality assignments. Figure 4 shows the images (and interface) for the case of preferring only ACCA. A comparison to Figure 2 shows how the improvement over cloud coverage (ACCA) does not seem to make up for the deterioration in seasonality quality (NDVI). Such a detrimental effect on NDVI, which was observed for all other factors as well, supports the argument that all factors must be considered simultaneously when building the "best" map. It is apparent from confronting the GMG problem that automated solvers are better suited to evaluate the effects of the interactions among multiple criteria (more than 2) than humans.

Conclusions

This paper has described an approach for generating high quality global image maps that incorporates an automated COP local search solver into a robust visual interface that allows users to manually manipulate solution images. Using the solver to manage the GLS2005 survey yielded measurable improvements in the quality of global image maps, with a beneficial reduction in the efforts required to produce those maps. GLS2005 scene selection was complicated by the utilization of multiple sensors, as well as requirements for gap-filling, which were not considerations in prior global map productions such as GeoCover-2000. A completely manual solution to optimal global map generation would have been infeasible. The combined approach we employed, using automation augmented by human guidance, proved to be a feasible compromise.

Based on a successful application of the GMG on the mid-decadal global Landsat data set, the GMG will be used for additional data set projects as well. One such mapping application has to do with the USGS goal to create a state mosaic for all 50 states in the US. The use of GMG has the potential of automating a large part of that effort. Due to the scanner failure on L7, GMG can greatly reduce the labor necessary to exploit the Landsat data archive. The L5 and L7 remote sensing spacecrafts are expected to continue service into 2012, GMG offers potential ongoing benefit with its ability to rapidly explore and assess large numbers of candidate solutions for regional and global Earth Science studies and other mapping applications. USGS and NASA are already planning for the next Decadal survey, GLS-2010. That plan includes leveraging on the success of LASSI and GMG.

The objective function criteria and aspects of the interface occasionally undergo refinements based on evolving customer requirements. For example, a criterion was recently added that considers whether a scene is predominantly agricultural. If so, L5 imagery is preferred because the artifacts resulting from L7 image pair compositing are more noticeable and problematic when gaps are filled through homogeneous farmland. Another recent change allows for more human intervention into the solution generated. For example, as the result of a recent update, if a user manually selects a L7 base-fill image pair, then the automated solver is not allowed to alter that selection, or reverse the base-fill images.

The global map generation problem provides an ideal domain for testing and evaluating constraint-based optimization solvers. Furthermore, the GMG solver is of significant potential benefit to the Earth Science research community, allowing scientists access to improved automated tools to study the Earth's changing eco-system. There are future plans to apply the approach described in this paper to generating complete moon maps using Clementine image data.

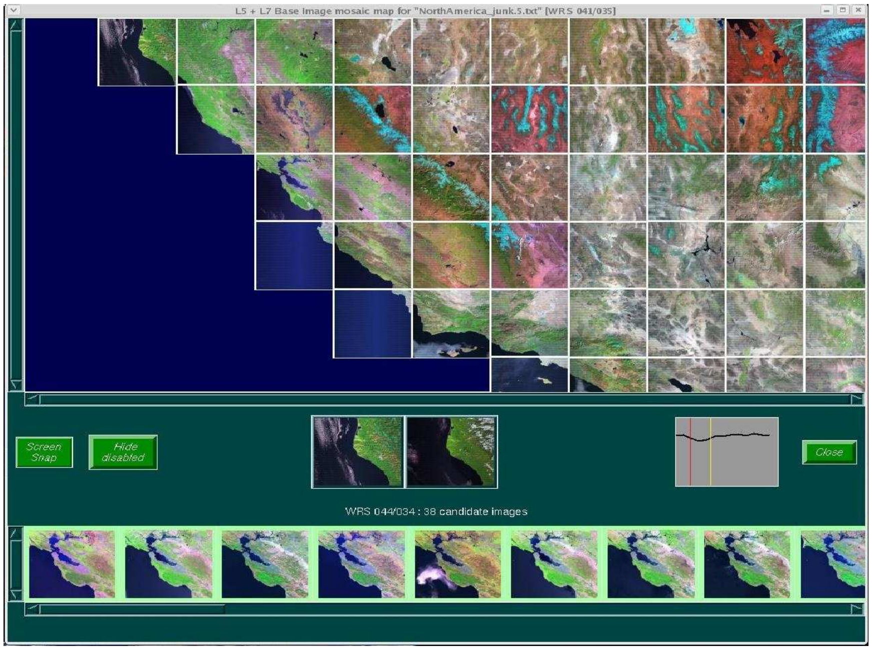

Figure 4: North America optimized for Cloud Cover only. LASSI Screen shot of thumbnail images of a GMG solution.

Acknowledgements

Thanks to Jeremy Frank and Steven Covington for their valuable contributions to this work.

References

Aarts, E., and Lenstra, J. K. 1997. *Local Search in Combinatorial Optimization*. John Wiley and Sons.

Cabon, B.; de Givry, S.; Lobjois, L.; Schiex, T.; and Warners, J. P. 1999. Radio link frequency assignment. *Constraints* 4(1):78–87.

Dechter, R. 2003. *Constraint Processing*. Morgan Kaufmann.

Franks, S.; Masek, J.; Headley, R.; Gasch, J.; Covington, S.; and Arvidson, T. 2008. *Large Area Scene Selection Interface (LASSI): Methodology of Selecting Landsat Imagery for the Global Land Survey 2005* . Submitted for Publication.

Hicks, I. V.; Koster, A.; and Kolotoglu, E. 2005. Branch and tree decomposition techniques for discrete optimization. *Tutorials in Operation Research, INFORMS* 1–29.

Khatib, L.; Gasch, J.; Morris, R.; and Covington, S. 2007. Local Search for Optimal Global Map Generation Using Mid-Decadal Landsat Images. *AAAI Workshop on Preference Handling for Artificial Inteligence*, 2007.

Larrosa, J., and Dechter, R. 2003. Boosting search with variable elimination in constraint optimization and constraint satisfaction problems. *Constraints* 8(3):303–326.

Larrosa, J.; Morancho, E.; and Niso, D. 2005. On the practical use of variable elimination in constraint optimization problems: 'still-life' as a case study. *JAIR* 23:421–440.

Larrosa, J. 2007. Javier Larrosa, personal correspondence.

Morris, R.; Khatib, L.; Gasch, J; and Covington, S. 2008. Automated Global Map Generation Using Mid-Decadal Landsat Images. *International Symposium on Aritificial Intelligence, Robotics and Automation in Space (iSAIRAS)*, 2008.

Samer, M., and Szeider, S. 2006. Constraint satisfaction with bounded treewidth revisited. In Benhamou, F., ed., *Principles and Practice of Constraint Programming - CP2006*, 499–513.