Compression of Digital Orthophotography Collections: Factors to Consider in the Compression of Large Data Sets of Geospatial Imagery

By Deette M. Lund

Illinois State Geological Survey
615 E. Peabody Drive
Champaign, IL 61820
Telephone: (217) 265-5642
Fax: (217) 333-2830
e-mail: lund@isgs.uiuc.edu

INTRODUCTION

In early 2006 the Illinois State Geological Survey (ISGS) was scheduled to receive several large digital orthophotography collections. As part of the contract requirements, the individual images were to be compressed and made available to the public via the Internet. In December 2005 and January 2006, research was conducted regarding compression formats, compression software, and target compression ratios. During the course of our research, it became clear that the technology and standards involved with the compression of geospatial imagery were fast changing.

Image compression techniques have improved in the past few years. It seemed that whichever compression format we chose, it would yield visibly better results than those available five or even two years ago. The visible results of image compression are only part of what needs to be considered when making decisions regarding the compression of geospatial imagery. Metadata is also very important to geospatial imagery. Information such as pixel size, geographic location, and the coordinate reference system are just a few of the critical pieces of metadata embedded in a compressed geospatial image file that GIS applications need to properly display the image. A metadata standard is necessary for the variety of compressed file formats to interact with the GIS applications appropriately. Metadata standards[1] for compressed geospatial imagery, in some compression file formats, were not fully established as of January 2006.

---

[1]By June 2006, the metadata standards issue involving JPEG 2000 had been resolved. According to the Open Geospatial Consortium the "GML in JPEG 2000 Inter-operability Experiment (GMLJP2)" initiative has been completed. Currently it appears that all of the phases have been fully developed. Time constraints have prohibited any further research into these latest developments. Further research will be needed to determine what that means.

FORMATS

Two popular compression formats were included in the research: JPEG 2000 (non proprietary), and LizardTech's MrSID. A third compression format, Earth Resource Mapper's ECW format, was included in the initial stage of our research but was excluded primarily to our long established relationship with LizardTech. The time constraints on our project did not allow time for building a new relationship with a different company.

JPEG 2000 Format

In 2004 & 2005 the JPEG 2000 compression format had become accepted as a standard by the International Organization for Standardization (ISO) and the International Electro-technical Commission (IEC). It became apparent during our study that JPEG 2000 was being developed in phases and that it was not fully developed (Morris, 2005). Several aspects of JPEG 2000 had been accepted as a standard by the ISO/IEC, but as of January, 2006 the geospatial aspects of the JPEG 2000 format were still in development and had not yet been approved as a standard. Another factor to consider with JPEG 2000 is that in 2003, according to Stuart Nixon, founder and CEO of Earth Resource Mapping (ER Mapper), there are at least three competing ways to store map projection information within a JPEG 2000 file, and our software developers use different methods (Thurston, 2003).

MrSID Format

LizardTech offers several compression algorithms within its latest upgrade of GeoExpress 6.0. Three that the ISGS considered were MrSID Generation 2 (MG2), MrSID Generation 3 (MG3), and JPEG 2000 (JP2). MG3 has improved compression capabilities. MG3 can compress

in lossless format, 2:1 for black and white imagery and up to 6:1 in color imagery (ratios will vary from image to image). The lossy compression for MG3 is also improved, generating up to 50% better compression ratios (depending on the image) than MG2. Unfortunately not all GIS software packages have caught up with the MrSID technology, including software developed by Environmental Systems Research Institute (ESRI). Some of ESRI's GIS software packages are still not fully capable of using the MG3 format and the majority of the GIS user community in Illinois uses ESRI technology. For this comparison only MG2 and JP2 were tested.

ECW Format

ER Mapper's ECW format was not considered for this comparison due to a number of factors. The first was due to time constraints on our project which did not allow time to establish a new relationship with a different company. Secondly there were patent litigation[2] issues at the time of our research. Earth Resource Mapping (ERM), the parent company of ER Mapper, was in litigation with Galdos, the parent company of LizardTech, over issues involving patent infringements (Thurston, 2003). LizardTech started the litigation and claimed their patent had been infringed. The companies have been in litigation since October 1999 and although it appeared it would be resolved soon, the ISGS could not wait for an outcome. A third factor included several documents available on the Internet that report comparisons between ECW and MrSID formats (GIS Services, 2005; Warmath, 2004). Those comparisons did not promote ECW as the better format. In contrast to those comparisons, we did get some positive feedback about ER Mapper and the ECW format from the Digital Mapping Techniques 2005 forum about image compression.

SOFTWARE

Two popular software packages were included in our research: Leica's Erdas Imagine, and LizardTech's GeoExpress. A third software package, GeoJasPer, was initially included in the research but was excluded early on in the research due primarily to its lack of technical support services.

---

[2]As of January 31, 2006 the litigation between Earth Resource Mapping and Galdos was settled (http://www.ermapper.com/company/news_view.aspx?PRESS_RELEASE_ID=398). Earth Resource Mapping won its claims against Galdos, but too late to be considered by the ISGS.

Leica's Erdas Imagine

Erdas Imagine provides free image compression within its software application. One limitation that is that it only provides compression for files up to 50 MB for the MrSID format files. The 2005 USGS NAPP-DOQ files exceed 170 MB in size and the 2005 USGS Urban Area files exceed 70 MB in size. The size of the files ruled it out as an option before another, not so obvious, factor came into play, which concerned the fact that Erdas used LizardTech's Software Development Kit (SDK) in setting up its compression capabilities. The developer has already made some encoding decisions for the user. Erdas only allows the user to change some of the multiple encoding options that are available with GeoExpress.

LizardTech's GeoExpress

Prior to December 2005, the ISGS had used LizardTech's MrSID Geospatial Encoder to compress all existing ISGS orthophotography collections. We needed to factor in the cost of an upgrade if we were going to use LizardTech's software again. The ISGS hadn't kept pace with LizardTech's software upgrades. This was primarily due to fiscal constraints and low usage of the software by staff after the initial purchase to compress the 1998-2000 NAPP DOQ collection. The upgrades at the ISGS had stopped just short of LizardTech's decision to use "data cartridges" (a file that keeps track of the amount of imagery that has been compressed) as its new way to charge customers for compression. LizardTech's new GeoExpress 6.0 would be able to compress imagery using either MrSID or JPEG 2000 formats and offered an unlimited "data cartridge" at a set price.

GeoJasPer

Before the ISGS started the actual compression tests of the two formats it was determined that the project team would need to use a software application that had a technical support system. Then, if there were trouble with the software itself or how it was handling compressions, the ISGS staff could use the support service to troubleshoot and fix any problems. Through this decision it was decided that the ISGS would not use GeoJasPer since there was no technical support system.

COMPARISON CHART

A comparison chart between the two major compression formats was developed in an effort to organize the

facts related to each factor in the decision-making process (Figure 1). This chart shows the factors an institution or agency should consider when making decisions about which compression format and software to choose. Some of the facts within the chart are time sensitive and may no longer be relevant.

End User

Another factor in our decision-making process that was not added to the chart was the end user. The ISGS had already "trained" its Clearinghouse user base to use MrSID compressed imagery. Using GeoExpress to com-

MrSID and JPEG 2000 Comparison*

FACTOR	MrSID (sid format)	JPEG 2000 (GeoJP2 format)

Software Choices	GeoExpress (LizardTech – PC, LINUX, SOLARIS* options)
	  * note - have experienced trouble with Solaris installation.
	    Also, instruction manual for command line encoding
	    could include better examples.
	Erdas (LizardTech SDK - only useful on files under 50 MB)
		GeoExpress (LizardTech)
		Erdas Imagine (software extension created from LizardTech
		    Software Development Kit (SDK))
		GeoJasPer (created from LizardTech SDK)
		ECW JPEG 2000 (ER Mapper)

Cost	GeoExpress 6.0 - Unlimited version = >$3000 or Data Cartridge Version = >$2000 per TB Erdas - different pricing available to each institution or agency	Erdas - different pricing available to each institution or agency GeoExpress - same as MrSID format costs GeoJasPer - free

Geography Markup Language (GML) standard	Follows the GML standard	Has GML in some cases. Still working on standardizations. Current
		    status of future standardization is not clear. Currently there are
		    at least 3 competing ways to store map projection information.

ISO Standard	No - because it is proprietary	Yes - but all phases not fully developed yet

ESRI Compatible	MrSID Generation 2 - Yes - but need to define projection or
	    provide an .AUX file
	MrSID Generation 3 - Not in all cases
		Yes - but potential issues with geospatial info
		  - depends on code writers choice of where to store geospatial
		    metadata (couldn't find any problems during limited testing)

Compatibility with other GIS Software Packages	MG2 - Majority of cases (with Plug-ins for a few) MG3 - Not in all cases http://www.gisservices.net/downloads/NYProgram.pdf (As of May 2004)	Not in all cases - and even then it might have problems with geospatial info
Web Browser (Free Viewer)	ExpressView Browser Plug-in	Yes - the ExpressView and a few others (some viewers are better than others)

Compatibility with Adobe CS2	Yes
	  - by using MrSID Decode (free)
	  - by using "Save as" in the ExpressView browser (be careful it
	    only saves the image visible on the screen at the time but it
	    will kick out a .TFW file if you save it to .TIFF format)
		Yes
		  - can place the image in Illustrator
		  - can also use same "save as" method described in sid format
		No - can not open in Photoshop

Compatibility with Other Graphics Software	Yes
	  - by using MrSID Decode (free)
	  - by using "Save as" in the ExpressView browser (be careful it
	    only saves the image visible on the screen at the time but it
	    will kick out a .TFW file if you save it to .TIFF)
		Many third party plug-in's available
		  - some are free
		  - some are free for the "lite" version and then you pay extra for
		    more bells and whistles

Generates log file (for metadata and statistics puposes)	GeoExpress
	    UNIX - Yes
	    PC - Yes
	Erdas - Yes
		Erdas - No
		GeoExpress - Yes
		GeoJasPer - Software not tested; No on-line information provided
		    about log file generation

Target -vs- Actual Compression Ratio	GeoExpress - can be much different 12:1 can result in 9.64:1
	Erdas - same as GeoExpress
		Erdas - No log file to list actual compression ratio information
		GeoExpress - stays more on target (from existing tests)
		    12:1 is 11.94:1
		GeoJasPer - Software not tested; No on-line information provided
		    about generating actual compression ratio information

Batch processing	UNIX - Yes
	PC
	    GUI - multiple file (not "true" batch processing)
	    CMD - batch processing (similar to UNIX - not tested)
		Erdas - possible according to help documents (not tested)
		GeoExpress
		    GUI - multiple file (not "true" batch processing)
		    CMD - batch processing
		GeoJasPer - possible according to on-line instructions (not tested)

Control over encode settings	UNIX - full (command driven)
	PC - full (can save established profiles), "pre-tuned" but user
	    can alter all settings
		Erdas - not as many options as GeoExpress
		GeoExpress - more control than Erdas, "not pre-tuned" like MrSID
		    (can be good or bad thing), can't control gamma or weight
		GeoJasPer - only controls target compression - no other settings

Generates world file	UNIX - Yes PC - Yes	Erdas - No instructions available about generating a world file GeoExpress - Yes GeoJasPer - No instructions available about generating a world file
Customizable Metadata	UNIX - Yes PC - Yes	Erdas - No GeoExpress - Yes GeoJasPer - No

*All costs and statistics current as of January 2006

Figure 1. Factors considered in comparison of MrSID and JPEG 2000 Compression formats.

press the images into either MrSID or JPEG 2000 format would result in no changes to user instructions and or viewer downloads. Researching how well other software would provide a compressed image that would be able to use the established viewer and whether the compressed images from that software would load properly into ArcSDE was beyond the time frame of the project. These considerations would need to be re-evaluated under different funding sources and time constraints.

CONCLUSION

In a technically challenging process, we considered the pros and cons of each compression format and each software option. We chose GeoExpress and the MrSID Generation 2 compression format for compressing the large data sets that would begin arriving in spring 2006.

The ISGS chose 8:1 for the target compression ratio for the 2005 USGS Chicago Urban Area color orthophotography collection. Differences between the original and the compressed imagery at actual size are not detectable. Zooming in beyond the reasonable usefulness of the image, at pixel level, the user can see a few changes. Those changes appear to be slight shifts in color on a few of the pixel groupings, but they are not easily detected. The average size of the uncompressed file is 71.5 MB. To keep the download time to a minimum, an 8:1 target compression ratio produced files under 10 MB in size. The average actual compression ratio for the 4527 files in this data set was 8.3:1.

The ISGS chose 10:1 for the target compression ratio for the 2005 USGS NAPP-DOQQ grayscale orthophotography collection. There are little-to-no differences between the original and the compressed imagery at

actual size. If users zoom in to 200%, "compression artifacts" (loss of edge detail and slight fuzziness) are visible. For the most part the "compression artifacts" in the compressed images do not affect the use of the images for research. The average size of uncompressed file is 177 MB (State Plane version). To keep the download time to a minimum, a 10:1 target compression ratio produced files around 20 MB in size. To date, we have compressed nearly one thousand of these State Plane version files. The average actual compression ratio for the ~900 State Plane version files that have been delivered is 9.6:1.

Due to rapid advances in standards and technology the facts are frequently changing in regards to image compression. Each institution or agency has its own particular factors to consider when dealing with image compression. The factors listed in Figure 1 should be used as a starting point or guide but the facts within the chart must be re-examined before deciding which formats and software to adopt.

REFERENCES

GIS Services, 2005, New York State Program - GeoExpress with MrSID: GIS Services, accessed at http://www.gisservices. net/downloads/NYProgram.pdf.

Morris, Steve, 2005, GML Content of JPEG 2000 format: North Carolina State University Library, December 30, 2005, personal communication (email).

Thurston, Jeff, 2003, Unlocking Data to Expanded Potentials: Interview with Stuart Nixon, Founder and CEO of Earth Resource Mapping, GEO Informatics Magazine, accessed at http://www.ermapper.com/document/doc.aspx?doc_ id=79.

Warmath, Eric, 2004, State Mapping Advisory Committee Meeting Notes: Report on Image File Compression Software, April 15, 2004, accessed at http://www.nbmg.unr.edu/smac/ apr2004.pdf

ADDITIONAL ONLINE RESOURCES

Brislawn, Christopher M., 2002, The FBI Fingerprint Image Compression Standard, accessed at http://www.c3.lanl. gov/~brislawn/FBI/FBI.html.

Digital Preservation Formats - http://www.digitalpreservation.gov/formats/index.shtml
   Jpeg 2000 - http://www.digitalpreservation.gov/formats/fdd/fdd000143.shtml
   Jpeg 2000 - http://www.digitalpreservation.gov/formats/fdd/fdd000140.shtml
   MrSID Generation 2 - http://www.digitalpreservation.gov/formats/fdd/fdd000031.shtml
   MrSID Generation 3 - http://www.digitalpreservation.gov/formats/fdd/fdd000184.shtml

Erdas Imagine Software Website
  http://gi.leica-geosystems.com/default.aspx

ER Mapper Software Website
  http://www.ermapper.com/

GeoJasPer Software Website (software function has changed since January 2006)
  http://www.dimin.net/software/geojasper/

GIS Monitor - Newsletter
  April 1, 2004
    LizardTech and Galdos Take on JPEG 2000.
    LizardTech/Earth Resources Mapping Lawsuit Judgment
  January 20, 2005
    LizardTech Introduces GeoExpress 5.0 with MrSID
  December 15, 2005
  Industry Survey: What was big news this year and what do you wish for next year?

Jakulin, Aleks, 2002-2004, Baseline JPEG and JPEG2000 Artifacts Illustrated, accessed at http://ai.fri.uni-lj.si/~aleks/jpeg/artifacts.htm.

Joint Photographic Experts Group - http://www.jpeg.org
  Jpeg 2000 - http://www.jpeg.org/jpeg2000/index.html

Library of Congress, How to View - The American Memory Collections, accessed at http://memory.loc.gov/ammem/help/view.html

LizardTech - Press Room
  March 23, 2004
  LizardTech, Galdos Systems Collaborate to Develop ISO Standard for JPEG 2000.
  May 3, 2004
    LizardTech Unveils MrSID Software Developer Kit with JPEG 2000.

LizardTech – GeoExpress Software
  http://www.lizardtech.com/

Morley, Karen, 2006, Avoid Pitfalls When Using JPEG 2000: GeoPlace.com, Tech Time Article, April 11, 2005, accessed at http://www.geoplace.com/uploads/FeatureArticle/0411tt.asp.

Pew Internet - American Life Project Demographics about internet use in America, accessed at http://www.pewinternet.org.

Wallace, Steve, 1999, Image Compression Software, accessed at http://www.directionsmag.com/features.php?feature_id=27.

Wikipedia
  Jpeg 2000 - http://en.wikipedia.org/wiki/Jpeg_2000
  MrSID - http://en.wikipedia.org/wiki/MrSID