Compression of Digital Orthophotography Collections

by Deette Lund
Illinois State Geological Survey

Portion of original Geotiff image tile shown at 50%. Calumet City, Cook County, Illinois.

Portion of original Geotiff DOQQ shown at 50%. Rock Island, Rock Island County, Illinois.

Factors to Consider in the Compression of Large Data Sets of Geospatial Imagery

A CHANCE TO CHANGE OUR COMPRESSION FORMAT

This information was initially gathered during December 2005 and January 2006 in an effort to determine the best compression format to use for the new orthophotography collections that were about to be delivered to the Illinois State Geological Survey (ISGS) for eventual web distribution. In deciding which compression format to use, another decision would need to be made about a software package. While pursuing this information one thing became clear. The compression of geospatial imagery is an ever changing world.

The good news is that compression technology appears to be improving, almost on a monthly basis. It appeared that whatever format was chosen it would yield visibly better results than for even two years ago. But don't forget to look beyond the visible results of compression and look at the image metadata. Metadata is 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 data that a GIS application would need to properly display a compressed geospatial image.

Prior to December 2005, the ISGS had used Lizardtech's MrSID Geospatial Encoder to compress all existing ISGS orthophotography collections. We needed to update our compression software because we 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 DOQQ collection. The upgrades at the ISGS had stopped short of Lizardtech's move to use "data cartridges" as it's new way to charge customers for compression, along with the license fees for their software application.

It was a challenge to navigate the pros and cons of each compression format and each software option. In the end we selected GeoExpress and the MrSID Generation 2 compression format for compressing the large data sets that were due to arrive in Spring 2006. So far the choice has been well received.

FACTORS CONSIDERED

● JPEG 2000 FORMAT & STANDARDIZATION
In 2004 & 2005 the JPEG compression format had become accepted as a standard by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). While examining the standardization, it was apparent that JPEG 2000 was being developed in phases and that it was not fully developed. Several of the phases had been accepted as a standard by the ISO/IEC, but as of January of 2006 the geospatial aspects of the JPEG 2000 format were still in development and had not yet been approved as a standard. Currently it appears that all of the phases have been fully developed. Time constraints have prohibited any further research into these latest developments. Another factor to consider with JPEG 2000 is that according to Stuart Nixon, founder and CEO of Earth Resource Mapping, there are at least three competing ways to store map projection information within a JPEG 2000 file depending on who the software developer is. This statement may not be entirely true at this printing. According to the Open Geospatial Consortium the "GML in JPEG 2000 Interoperability Experiment (GMLJP2)" initiative has been completed. Further research will be needed to determine what that means.

● ECW FORMAT
ER Mappers ECW format was not considered for this comparison due to a number of factors, including the fact that there were several documents available that reported the comparisons between ECW and MrSID formats. The comparisons don't promote ECW as the better format. We did however get feedback from Bob Parker of the U.S. Department about ER Mapper and the ECW format. Another factor being that Earth Resource Mapping, the parent company of ER Mapper, is still in litigation with Gabbs, the parent company of Lizardtech, over issues involving each companies compression software. The final factor was that the ISGS already had an established relationship with Lizardtech.

● MrSID FORMAT
Lizardtech offers several ways of compression within its latest upgrade of GeoExpress 6.0. These 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 packages have caught up with the MrSID technology, including ESRI. Some of ESRI's software is still not fully capable of using the MG3 format. For this comparison only MG2 and JP2 were tested.

● SUPPORT SERVICES
Before the ISGS started the actual compression tests of the two formats it was determined that the survey would need to use a software that had a support system behind it. 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 GeoasPer since it did not have those services in place.

● SOFTWARE WITH SUPPORT SERVICES & NO "DATA CARTRIDGE
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 sid format files. The 2005 USGS NAPP-DOQQ files exceed 170 MB in size and the 2005 USGS Urban Area files exceed 70 MB in size. The file size for these 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 time encoding decisions for the user. Erdas only allows the user to change some of the multiple encoding options that you would be given in GeoExpress.

● KEEPING CLEARINGHOUSE USERS HAPPY
Another thing to consider in the decision making process was the end user. The ISGS had already "trained" its Clearinghouse user base to use MrSID compressed imagery. Using GeoExpress to compress 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 reviewed under different funding sources.

MrSID and JPEG2000 Comparison*

FACTOR	MrSID (sid format)	JPEG 2000 (GeoJP2 format)
Software Choices	GeoExpress (PC, UNIX, SOLARIS* options) * note - have experienced trouble with Solaris installation. Also, instruction manual for command line encoding could use better examples. Erdas (up to 50MB files)	GeoExpress (Lizardtech) Erdas Imagine (made from Lizardtech Software Development Kit (SDK)) GeoasPer (made from Lizardtech SDK) ECW JPEG 2000 (ER Mapper)
Cost	GeoExpress 6.0 - Unlimited version = >$3000 - or Data Cartridge Version = >$2000 per TB Erdas - ??	Erdas- ?? GeoExpress - same as sid format costs GeoasPer - Free
Geography Markup Language (GML) standard	Follows the GML standard	Has GML in some cases. Still working on getting things standardized. Current status of when it will be standardized is not clear. One companies contention that there are currently at least 3 competing ways to store map projection information.
ISO Standard	Yes - 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 (although I couldn't find any problems with my limited testing)
Compatibility with other GIS Software Packages	MG2 - majority (with Plug-ins for a few) MG3 - Not in all cases	Not in all cases - and even then it might have problems with geospatial info
Web Browser (Free Viewer)	ram.pdf (2004 list) ExpressView Browser Plugin	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 the file if you save it to tif)	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/ Publication 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 the file if you save it to tif)	lots of third party plug-in's out there - some are free - some are free for the 'lite' version and then you pay extra for them bells and whistles
Creates log file for process statistics such as compression ratios	GeoExpress UNIX - Yes PC - Yes Erdas - No	Erdas - No GeoExpress - Yes GeoasPer - Don't know
Hitting the "Target" - target -vs- actual compression ratio	GeoExpress - can be much different 12:1 can result in 9.64:1 Erdas - same as GeoExpress	Erdas - Don't know it doesn't give you the info GeoExpress - stays more on target (from existing tests) 12:1 is 11.94:1 GeoasPer - Don't know
Batch processing	UNIX - Yes PC - GUI - multiple file (not "true" batch processing) - CMD - batch processing (similar to UNIX - not tested)	Erdas - possible (not tested) GeoExpress - CMD - batch processing GeoasPer - yes (not tested)
Control over encode settings	UNIX - full (command driven) PC - full (can see established presets), 'pre-tuned' but user can alter all settings	Erdas - not as many options as GeoExpress GeoExpress - more control than Erdas, 'not pretuned' like MrSID (can be good or bad thing), can't control gamma or weight GeoasPer - only controls target compression - no other settings
Generates world file	UNIX - Yes PC - Yes	Erdas - Don't know GeoExpress - Yes GeoasPer - Don't know
Customizable Metadata	UNIX - Yes PC - Yes	Erdas - No GeoExpress - Yes GeoasPer - No

*All costs and statistics current as of January 2006

Target vs. Actual Compression Ratios*

In practice, the algorithm invariably yields a different actual compression ratio than the target because of unpredictable amounts of lossless entropy coding gain, principally due to variable amounts of blank space in the images. The exact compression ratio achieved will vary from image to image. For this reason, the target compression ratio should be thought of as setting the distortion level rather than the precise compression ratio for the compressed images.

Lizardtech recommends not compressing beyond the following target compression ratios – 20:1 for color (RGB) and 12:1 for grayscale imagery.

		Original File Size (MB)	Compressed File Size (MB)	Target Compression Ratio	Actual Compression Ratio
SID FORMAT	Color Imagery (RGB)	71.5	13.8	5:1	5.20:1
		71.5	7.0	10:1	10.27:1
		71.5	4.7	15:1	15.33:1
		71.5	3.6	20:1	19.76:1
	Grayscale Imagery	177.0	34.2	5:1	5.17:1
		177.0	17.9	10:1	9.91:1
		177.0	11.6	15:1	15.31:1
		177.0	8.2	20:1	21.53:1
JP2 FORMAT Can not adjust gamma or weight	Color Imagery (RGB)	71.5	14.3	5:1	4.99:1
		71.5	7.2	10:1	9.97:1
		71.5	4.8	15:1	14.94:1
		71.5	3.6	20:1	19.90:1
	Grayscale	177.0	35.4	5:1	4.99:1
		177.0	17.7	10:1	9.98:1
		177.0	11.8	15:1	14.95:1
		177.0	8.9	20:1	19.91:1

*This data represents Lizardtech GeoExpress compression results. GeoExpress Command Line, Unlimited - Version 6.0.0.1331.Bld, 1304. Jar

Resources

Web Resources:
Brisben, Christopher M.
  The FBI Fingerprint Image Compression Standard.
  http://www.c3.lanl.gov/~brislawn/FBI/FBI.html
Digital Formats
  http://www.digitalpreservation.gov/formats/index.shtml
  http://www.digitalpreservation.gov/formats/fdd/fdd000143.shtml
JPEG 2000
  Image Compression: JPEG 2000, image formats/fdd/fdd000147.shtml
  MrSID Generation 1. http://www.digitalpreservation.gov/formats/fdd/fdd000161.shtml
  MrSID Generation 2. http://www.digitalpreservation.gov/formats/fdd/fdd000166.shtml
GeoExpress
  Interview with Stuart Nixon, founder and CEO of Earth Resources Mapping
GIS Services 2005
  New York State Program - GeoExpress with MrSID
  http://www.gisservices.net/downloads/NYProgram.pdf
  MrSID Generation 2
  Baseline JPEG and JPEG2000 Artifacts Illustrated. http://a.libri.p.liceti.net/~aklis/gen/artifacts.htm
Joint Photographic Experts Group
  http://www.jpeg.org/
  http://www.jpeg.org/jpeg2000/index.html
Library of Congress
  How to View - The American Memory Collections.
  http://memory.loc.gov/ammem/help/view.html
Mersky, Karen 2004
  MrSID Users Using JPEG 2000 GeoExpress, Tech Time Article, April 11, 2005.
  http://www.geophoto.com/uploads/FeatureArticle041106.asp/
Pew Internet - American Life Project
  Demographics about internet use in America. http://www.pewinternet.org/

Web Resources (cont):
Wallace, Steve 1999
  Image Compression Software. http://www.directionmag.com/features.php/feature_id=27
  March 20, 2004
  State Mapping Advisory Committee Meeting Notes. Report on Image File Compression
  Kruger, Roger. March 11, 2004. http://www.mn.gov/mn/us/smac/apr2004.pdf
Wikipedia
  Jpeg 2000. http://en.wikipedia.org/wiki/JPEG
  MrSID. http://en.wikipedia.org/wiki/MrSID

GIS Monitor - Newsletter
  April 1, 2004
  New MrSID version debuts. JPEG 2000,
  Lizardtech/Earth Resources Mapping Lawsuit Judgement
  January 20, 2005
  Lizardtech introduces GeoExpress 5.0 with MrSID
  December 15, 2005
  Industry Experts What was big news this year and what do you wish for next year?
Lizardtech - Press Room
  March 23, 2004
  Lizardtech , Galileo Systems Collaborate to Develop ISO Standard for JPEG 2000
  May 2, 2004
  About Polish: When Using JPEG 2000 GeoPlanet.com. Tech Time Article, April 11, 2005.

Personal Correspondence:
Morris, Steve (North Carolina State University Library)
  December 30, 2005. Email Subject: GML Content of JPEG 2000 Review

---

Comparison of Target Compression Ratios (Color)

SID* | Geotiff (original uncompressed image for comparison) | GeoJP2

5:1 — There is no detectable change between the original image and the two compressed formats at this target compression ratio.

10:1 — If you look at the diagonal lines in the parking lot you will start to see some blurring occuring. More in the GeoJP2 than in the SID image.

15:1 — More compression artifacts are starting to show up in both compressed images. The compressed images are blurring more but are still usable.

20:1 — Note the appearance of "blocking artifacts" (small square blocks all over the image). These are caused by the methods used for wavelet compression.

Shown at 200%

2005 USGS Chicago Urban Area Digital Orthophotography

Area of Coverage: Entirety of Cook, DuPage, Kane, Lake, McHenry and Will counties.
Imagery Source: True (natural) color film. Acquired during the leaf-off period from March 1-April 30, 2005.
Projections & Datum: Transverse Mercator, UTM grid zones 15 and 16, NAD 83 datum.
Spatial Resolution: 0.3 x 0.3 meter ground sample distance per pixel for the UTM based orthophotography.
Data Format: Uncompressed GeoTIFF with accompanying FGDC-compliant metadata.
Data Files: Each orthorectified image tile represents a 1,500 x 1,500 meter ground area produced at even 1,500 meter grid lines, with no overlap between image tiles. Corner coordinates are based on the UTM grid. Number of image files is 4527 and file size is approximately 73 Mb for each image tile.

Color Compression Decision

SID* | Geotiff (original uncompressed image for comparison) | GeoJP2

Urban
Agricultural
Forested
Industrial

Shown at 100%
Target Compression Ratio (sid and jp2 format) = 8:1

The ISGS chose 8:1 for the target compression ratio for the 2005 USGS Chicago Urban Area color orthophotography collection. As you can see by the images to the left, there is little, if any, difference between the original and the compressed imagery, in either compression format. Only if you zoom in beyond the reasonable usefulness of the image, at pixel level, do you see any changes. Those changes appear to be slight changes in color on a few of the of the pixel groupings and are not easily detected without a lot of concentration. 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 this data set was 8.3:1 for 4527 files.

Grayscale Compression Decision

SID* | Geotiff (original uncompressed image for comparison) | GeoJP2

Urban
Agricultural
Forested
Industrial

Shown at 100%
Target Compression Ratio (sid and jp2 format) = 10:1

The ISGS chose 10:1 for the target compression ratio for the 2005 USGS NAPP-DOQQ grayscale orthophotography collection. As you can see by the images to the left, there are little to no differences between the original and the compressed imagery at actual size, in both compression formats. If you zoom in to 200% you can start to see "compression artifacts" in both compression formats. For the most part the "compression artifacts" in the compressed images do not affect the use of the images for research. To keep the download time to a minimum, an 10:1 target compression ratio produced files around 20 MB in size. So far the State Plane version of the data are the only files that have had close to a thousand files compressed. The average actual compression ratio for the State Plane portion of the data set that has been delivered is 9.6:1.

2005 USGS Illinois NAPP - DOQQ Digital Orthophotography

Area of Coverage: Entire 96-county Illinois area exclusive of Cook, DuPage, Kane, Lake, McHenry and Will counties.
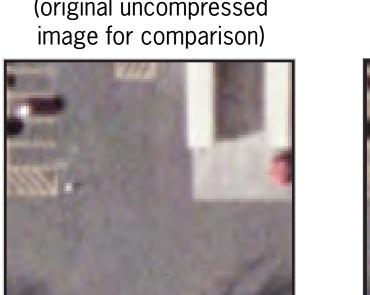Imagery Source: Black-and-white film. Acquired during the leaf-off period from February 15-April 30, 2005.
Projections & Datum: Transverse Mercator, UTM grid zones 15 and 16, NAD 83 datum.
Spatial Resolution: 0.5 x 0.5 meter ground sample distance per pixel for the UTM based orthophotography.
Data Format: Uncompressed GeoTIFF with accompanying FGDC-compliant metadata.
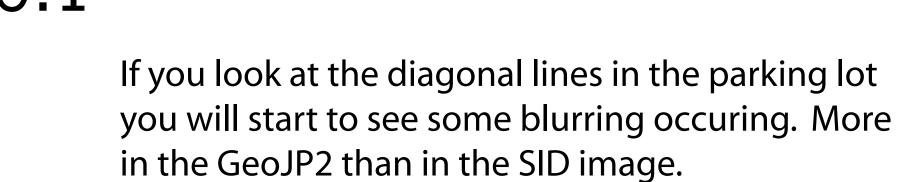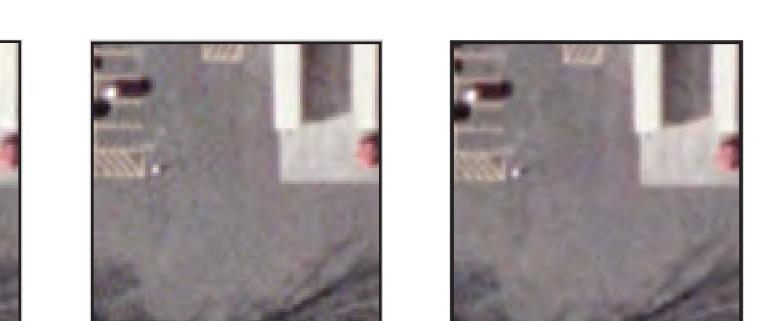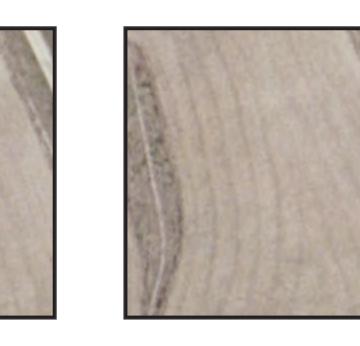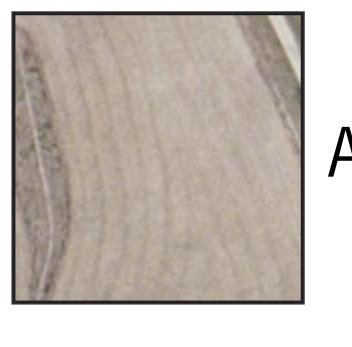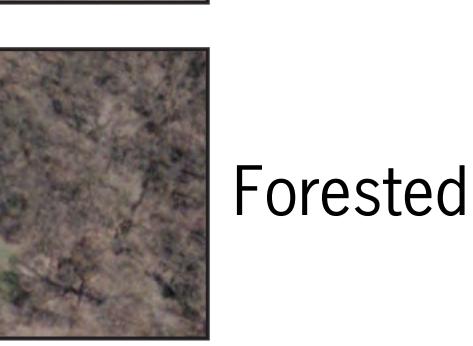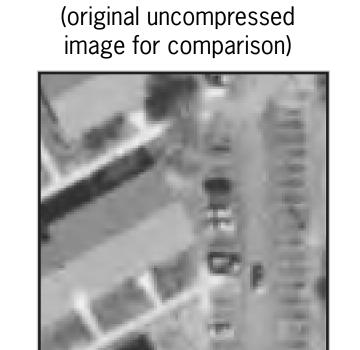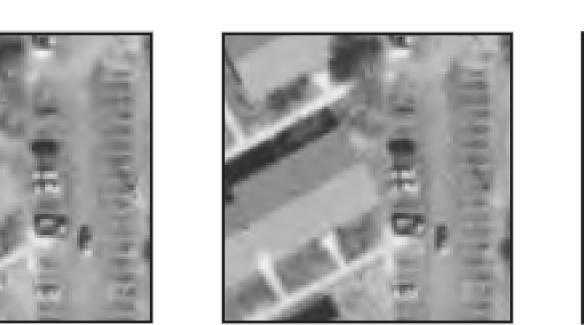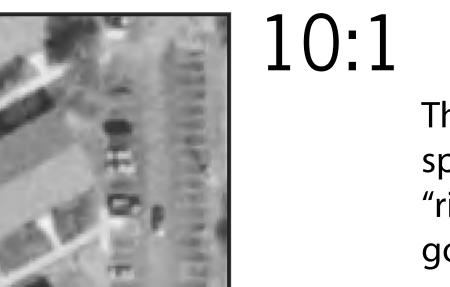Data Files: Each image file conforms to the USGS 3.75' x 3.75' Digital Orthophoto Quarter Quadrangle (DOQQ) standard. Estimated number of DOQQ image files is 3,262 and file size is approximately 1.75 Mb for each DOQQ.

Comparison of Target Compression Ratios (Grayscale)

SID* | Geotiff (original uncompressed image for comparison) | GeoJP2

5:1 — A change is already detectable in the GeoJP2 image. Two horizontal "ridges" appear to be running across the images.
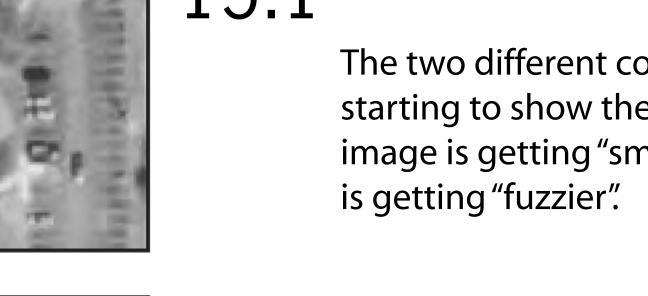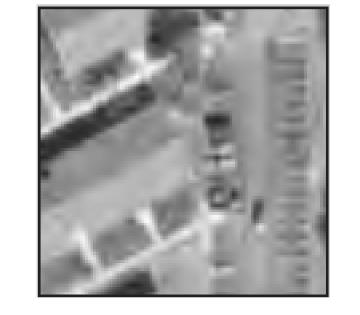
10:1 — The sharp edges of the roof lines and parking spaces are already starting to blur. But the "ridges" (an unpredictable artifact) appear to be gone in the GeoJP2 image.
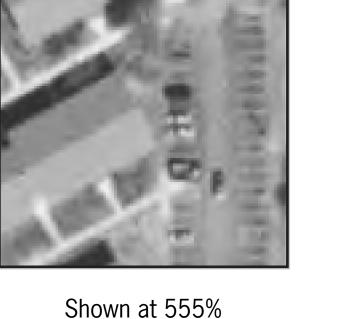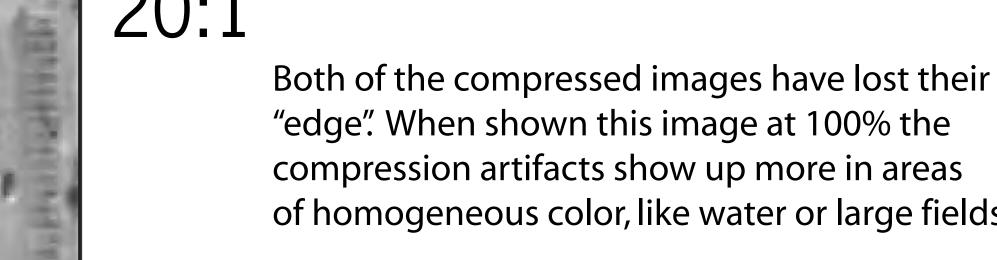
15:1 — The two different compression algorithms are starting to show their differences. The GeoJP2 image is getting "smoother," while the SID image is getting "fuzzier."

20:1 — Both of the compressed images have lost their "edge". When shown this image at 100% the compression artifacts show up more in areas of homogeneous color, like water or large fields.

Shown at 555%

*All sid format compressed images were compressed as MrSID Generation 2.