A reprint from

OPTICAL ENGINEERING

ISSN 0091-3286

Evaluation of the tactical utility of compressed imagery

John M. Irvine, MEMBER SPIE
Science Applications International Corporation
20 Burlington Mall Road, Suite 130
Burlington, Massachusetts 01803
E-mail: john.m.irvine@saic.com

Barbara A. Eckstein
McClendon Automation
14701 Willard Road
Chantilly, Virginia 20151

Robert A. Hummel
Defense Advanced Research Projects Agency ISO
3701 N. Fairfax Drive
Arlington, Virginia 22203

Richard J. Peters
Science Applications International Corporation
4001 North Fairfax Drive, Suite 300
Arlington, Virginia 22203

Rhonda Ritzel
Science Applications International Corporation
McLean/Enterprise Center
8301 Greensboro Drive
P.O. Box 50132
McLean, Virginia 22102

Abstract. The effects of compression on image utility are assessed based on manual exploitation performed by military imagery analysts (IAs). The original, uncompressed synthetic aperture radar imagery and compressed products are rated for the Radar National Imagery Interpretability Rating Scale (NIIRS), image features and sensor artifacts, and target detection and recognition. Images were compressed via standard JPEG compression, single-scale intelligent bandwidth compression (IBC), and wavelet/trellis-coded quantization (W/TCQ) at 50-to-1 and 100-to-1 ratios. We find that the utility of the compressed imagery differs only slightly from the uncompressed imagery, with the exception of the JPEG products. Otherwise, both the 50-to-1 and 100-to-1 compressed imagery appear similar in terms of image quality. Radar NIIRS indicates that even 100-to-1 compression using IBC or W/TCQ has minimal impact on imagery intelligence value. A slight loss in performance occurs for vehicle counting and identification tasks. These findings suggest that both single-scale IBC and W/TCQ compression techniques have matured to a point that they could provide value to the tactical user. Additional assessments may verify the practical limits of compression for synthetic aperture radar (SAR) data and address the transition to a field environment.
[DOI: 10.1117/1.1475740]

Subject terms: image compression; Joint Photographic Experts Group compression; intelligent bandwidth compression; wavelet/trellis-coded quantization; imagery analyst.

Paper 010135 received Apr. 13, 2001; revised manuscript received Sep. 27, 2001; accepted for publication Dec. 10, 2001.

1 Introduction

The utility of compressed imagery has become increasingly important as improvements in sensor technology yield imagery at data rates that far surpass the bandwidth of most available non-line-of-sight communication channels. The Defense Advanced Research Projects Agency (DARPA) has developed compression algorithms, such as intelligent bandwidth compression (IBC), to compress synthetic aperture radar (SAR) imagery. The compressed imagery must fit the bandwidths of one or a few T1 data links provided by commercial satellite communications while preserving the fidelity of imagery as required by tactical users. Recent studies have evaluated the impact of compression on the human imagery analyst (IA) in the exploitation of nonliteral, digital SAR imagery.[1]

This study evaluated the utility of compressed imagery by addressing two aspects of the compressed image product:

1. changes in image quality associated with image compression
2. differences in IA performance associated with compression

The IAs evaluated several versions of each image, including the original, uncompressed image. Images were compressed using the Joint Photographic Experts Group (JPEG) algorithm, the IBC algorithm, and the wavelet/trellis-coded quantization (W/TCQ) algorithm.

2 Approach

The impact of compression on image quality and utility was assessed comparatively using several renditions of the test imagery. Performance metrics quantified image quality and IA performance differences as a function of image compression. Image quality was graded using the Radar National Imagery Interpretability Rating Scale (NIIRS), as well as subjective NIIRS-like ratings for image features such as sidelobes, shadows, contrast, etc. Image utility was determined by the interpretability of context information and of target information, and was quantified by the accuracy of vehicle counting, vehicle classification and identification, and identification of military activity.

3 Imagery Requirements

3.1 *Image Selection*

The robustness of any evaluation depends on the breadth and size of the imagery sample set in the assessment. The imagery should span a range of conditions and collection

 0091-3286/2002/$15.00

Table 1 Imagery selected for this evaluation, a mixture of desert and temperate environments and various grazing angles.

Image Source	Grazing Angle (deg)	Number of Images
Desert environment	10–15	5
Temperate environment	10–15	3
Temperate environment	15–45	4
Total		12

parameters, including resolution (i.e., impulse response or IPR), grazing angle, scene content, and target densities, with multiple background types and a variety of target orientation, terrain, vegetation conditions, and cultural clutter. To capitalize on existing data that span a range of conditions, this study used SAR imagery from a desert environment with relatively benign clutter, and from a temperate climate with higher natural clutter.

3.2 *Sample Size*

The total number of scenes must balance the trade-off between the desired level of precision for a statistical analysis and the amount of time required to complete the evaluation. Statistical power calculations and historical experience drove the decision to use 12 original scenes. Generating the various compression products yielded 72 images that each IA viewed during the evaluation (Table 1). Each IA trained on a 13th image.

3.3 *Image Truth*

The subjective image quality questions, such as NIIRS ratings, cannot be quantified against "image truth." Image truth relates only to the objective image questions, such as vehicle counts, vehicle classification and identification, background identification, and identification of military activity. Wherever possible, ground truth was extracted from the record for the image collections. When certain objects in the image were not mentioned in the ground truth record, then an expert IA determined the correct answers, using available collateral data.

4 Image Compression Products

This study evaluated imagery compressed by 50-to-1 and 100-to-1 (Table 2). The IBC, W/TCQ, and JPEG algorithms, all of which are lossy, were applied to each scene, producing three versions of 50-to-1 compression. The IBC and W/TCQ algorithms also were applied for 100-to-1 compressed products. JPEG was not used for 100-to-1 compression because of the obviously poor quality of the result. Including the original, uncompressed image, each IA viewed six renditions of the same scene. Figure 1 shows a sample image and a target chip, i.e., a subset of the full image.

Table 2 Compression algorithms and ratios of compression.

Compression Algorithm	Compression Ratio
None	1:1
JPEG	50:1
W/TCQ	50:1
W/TCQ	100:1
IBC	50:1
IBC	100:1

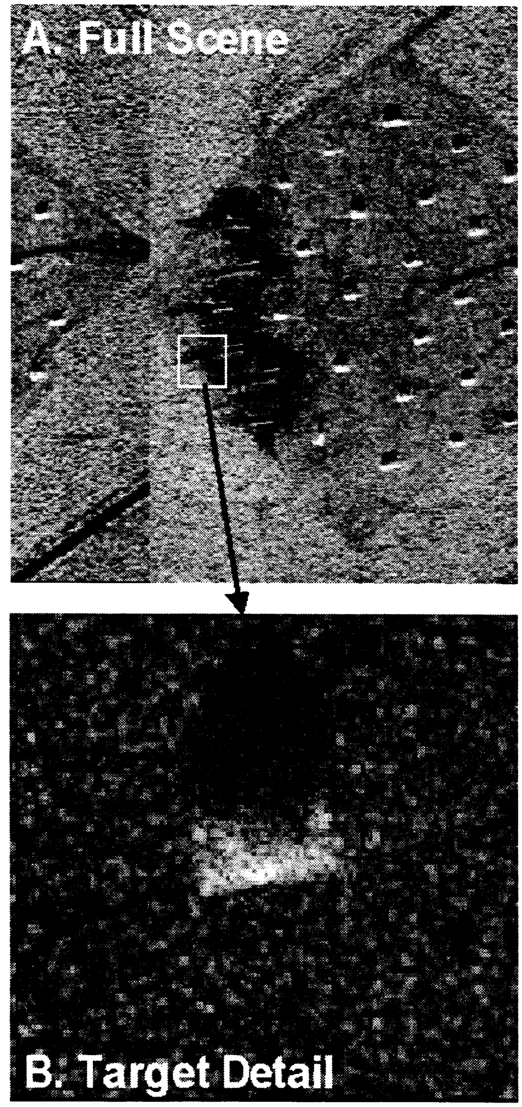

Fig. 1 One of the uncompressed SAR images evaluated by imagery analysts: (a) full scene showing multiple targets and several revetments at the left side of the image and (b) example target chip, showing the detail of the bright target and its shadow.

The three image compression algorithms differ with respect to the type of data they process and, consequently, are applied at different stages in the image chain. In all cases, the starting point is the 32-bit complex radar image and the product displayed to the IA was an 8-bit detected image. IBC operates directly on the complex data to produce a compressed product. Following the decompression, the data are mapped to an 8-bit detected product for display. For the W/TCQ algorithm, the first step is the formation of a 16-bit magnitude image, which W/TCQ compresses. Again, following decompression the image is mapped to 8 bits for display. The JPEG method employed in this study operated

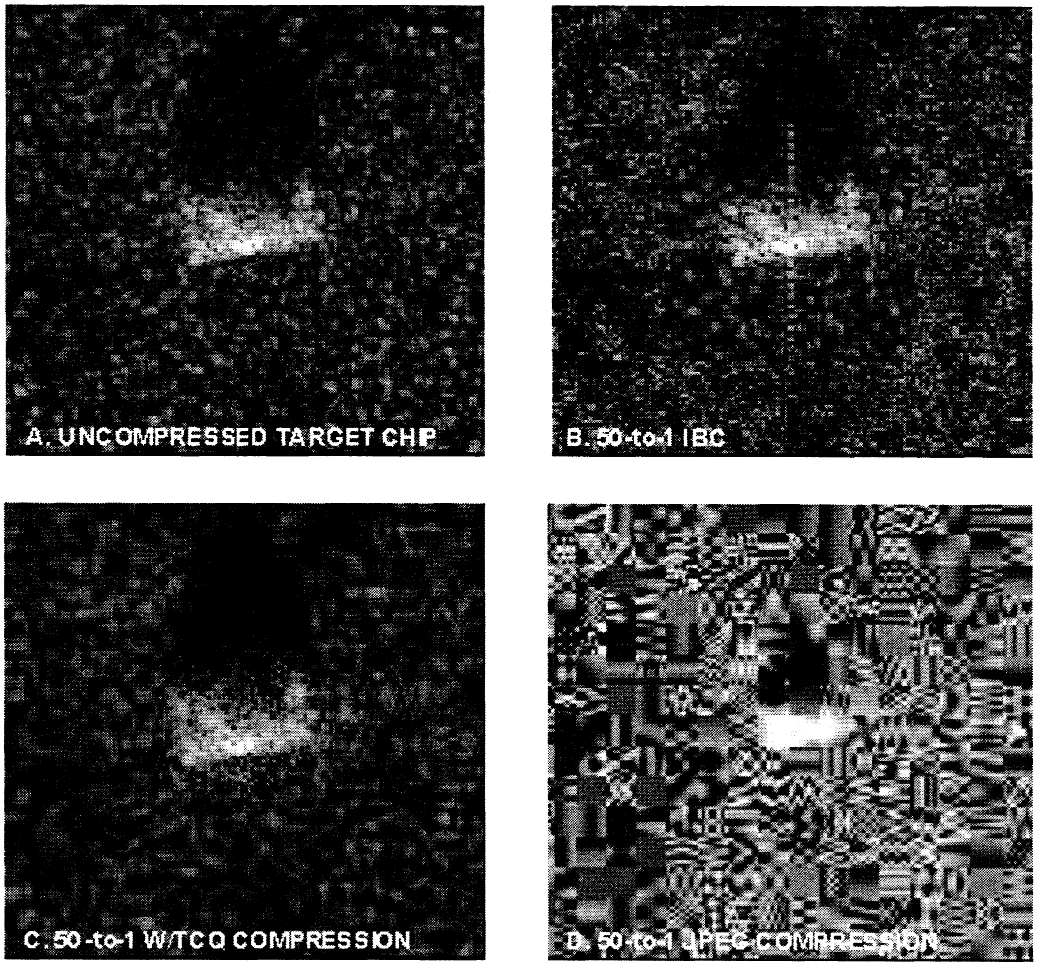

Fig. 2 Target chip from Fig. 1 before and after 50-to-1 compression, showing image compression artifacts from the various compression algorithms: (a) original, uncompressed target chip; (b) target chip after 50-to-1 IBC; (c) target chip after 50-to-1 W/TCQ compression; and (d) target chip after 50-to-1 JPEG compression.

on 8-bit data, so the conversion from complex to detected data and the mapping to 8-bits occurred prior to compression.

4.1 *IBC Products*

In IBC, a focus of attention algorithm identifies image regions that possibly contain militarily significant targets and labels them as regions of interest.[1–5] Small regions of pixels including the targets are extracted from the total image and are minimally compressed by quantizing the Fourier coefficients to yield modest data compression over the target region. Speckle noise is filtered from the remaining background image. The resulting smoothed image is then highly compressed using a wavelet-based technique. The result is compression ratios potentially exceeding 100-to-1 while preserving at relatively high fidelity the targets of interest and providing the scene context via the compressed background regions.

Figure 2(b) shows a target chip after 50-to-1 IBC compression. Comparison to Fig. 2(a) (the uncompressed target chip) indicates that compression has increased the ringing on the bright area of the target in this example. IBC also altered the speckle properties of the image.

4.2 *W/TCQ Compression*

The W/TCQ technique uses a combination of wavelet compression with trellis-coded quantization. Wavelet compression breaks down the image into many interrelated smaller components that are ranked in a hierarchy, so that each level contains more details than the next.[6–8] The hierarchy is a pyramid of repetitive application of frequency scaling to wavelet functions. The high-frequency bands contain the detailed information.

Trellis-coded quantization establishes a book of codewords for a given image.[7,9] The codebook is divided into four subsets using the criterion that the Euclidean distance between codewords in a subset is minimized. The image is then encoded using the codewords and trellis path (i.e., codeword sequence) that minimize the cumulative distortion of the encoded product.

Figure 2(c) depicts the target chip after 50-to-1 W/TCQ compression, showing that the image has lost detail. Edges are not as sharp, and the background variance has decreased and become smoother. Also, the speckle on the object itself has increased, although the W/TCQ processing smoothes the background speckle.

4.3 *JPEG Compression*

This study used the discrete cosine transform (DCT) JPEG algorithm to compress imagery by 50-to-1. JPEG divides[6] the 8-bit image into 8×8 minimum coding units or neighborhoods, and computes the forward discrete cosine transform (FDCT) of each neighborhood. The quantizer rounds off or smoothes the FDCT coefficients, and then encodes the result. For decompression, JPEG recovers the quantized FDCT coefficients from the compressed data stream, takes the inverse transform, and displays the image. The image is then upsampled using interpolation to restore the image to its original size.

Figure 2(d) presents the product of 50-to-1 DS-DCT-JPEG compression. The zoomed object in this compression product is smaller and has a somewhat different morphology compared to the original, uncompressed image. Additionally, the nature of the background has changed. The neighborhoods or coding units are visible, as evidenced by the patchwork pattern in the background region.

5 Image Evaluation Metrics

5.1 *Image Quality Evaluation Criteria*

Image quality metrics were designed to assess SAR-specific image artifacts and features. Thus each IA rated the radar NIIRS, as well as shadows, contrast, sidelobes, speckle, and motion blurring. The same image quality questions were asked for all 12 images.

5.1.1 *Radar NIIRS*

A primary measure of image interpretability is the Radar NIIRS, a task-based scale that standardizes the quantification of the interpretability or potential usefulness of imagery.[10–13] This approach to quantifying image quality is necessary because of the inability of simple physical image parameters, such as resolution, to adequately predict image interpretability.

The NIIRS consists of 10 graduated levels, from 0 to 9, with several interpretation tasks or criteria defining each level. Radar NIIRS level 1 imagery, for example, can be used to detect the presence of aircraft dispersal parking areas, a large cleared swath in a densely wooded area, the presence of piers and warehouses (at a port facility), and the presence of road or railway lines of communication (LOCs). Some of the tasks possible with high quality NIIRS level 8 imagery are the identification of an SA-6 transloader when other SA-6 equipment is present or differentiation between the fuselages of a HIND and HIP helicopter. Appendix A describes the Radar NIIRS criteria for levels 0 through 9. The IAs rated the Radar NIIRS of each image.

5.1.2 *Ratings of SAR image characteristics*

The IAs rated each image on SAR characteristics using nine-point subjective scales, similar to the radar NIIRS approach. The definitions for the extrema of these scales were:

1. shadows: 1, shadows not visible; 9, clear crisp shadows clearly showing target information
2. contrast: 1, scene is severely "washed out;" 9, excellent contrast, clearly showing image features
3. sidelobes: 1, extreme sidelobes or blooming obscures the targets; 9, minimal sidelobe evidence; target details are clear
4. speckle: 1, severe speckle; no distinction among background types; 9, no speckle problems, backgrounds clearly distinguishable
5. blurring: 1, severe blurring renders scene unexploitable; 9, no evidence of blurring, scene is clearly focused

5.2 *Image Exploitation Evaluation Criteria*

The ability of the IAs to detect, classify, and identify important targets and features established the utility of the imagery. These objective measures of image utility were quantified using ground truth information. For example, each IA was asked to count the vehicles in a region of the image, determine whether an object is a wheeled or tracked vehicle, and identify a vehicle.

Subjective questions assessed the ability of an IA to extract information related to certain essential elements of information (EEI). These questions included identifying the military activity in the scene, rating background quality and LOC extraction on a NIIRS-like scale, and identifying features in the background.

The specific exploitation questions differed for each of the 12 images, based on scene content. Every image included one question on LOC extraction confidence and at least one question on background quality. Certain images had two and three background quality and identification questions. Some scenes did not ask for a vehicle count, whereas others requested vehicle counts in one or two boxes indicated on the image. Similarly, the number of questions on vehicle classification and identification varied with image, and only a few images required the IAs to identify military activity.

6 Image Evaluation Procedure

6.1 *Software Tools*

This evaluation used the VITec Electronic Light Table (ELT) software hosted on Sun workstations. Figure 3 shows the VITec screen display for one of the images. The questions on image quality, i.e., rating Radar NIIRS, shadows, contrast, sidelobes, speckle, and blurring from 1 to 9, applied to the entire image. Similarly, the LOC confidence question applied to the entire image. Most image utility questions, however, were asked for designated regions within the image. The boxes and circles on the image delineated the regions for these specific questions. For example, boxes 1 and 2 correspond to the two questions on the background quality and identification. Circles 3, 4, and 5 correspond to vehicle classification and identification questions.

6.2 *Image Evaluation Process*

The evaluation process began each morning with a briefing to the IAs explaining the overall purpose of the assessment, orienting them to the VITec ELT program, and explaining the types of questions that would be asked. Each IA re-

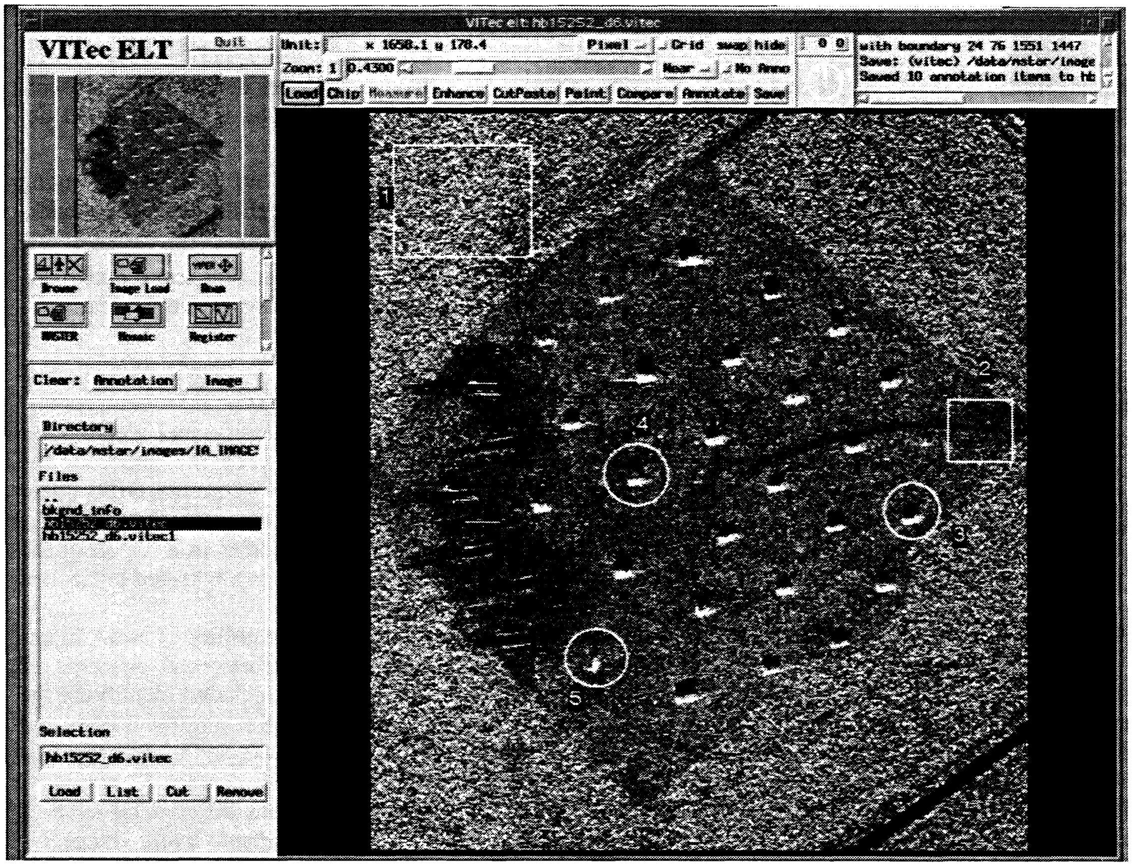

Fig. 3 Typical VITec ELT screen display of an image used in this evaluation, showing the numbered and circled targets that the imagery analysts were asked to identify.

ceived a hardcopy packet of image evaluation questionnaires that had been sorted for the correct order for viewing the imagery.

The IAs then viewed the images in the pseudorandomized order, answering the questions related to each image on the hardcopy form. They had no *a priori* information on whether a particular rendition of a scene was compressed.

The IAs were allowed as much time as they desired for each image. Once all of the images had been viewed, the IAs completed a short questionnaire on their background and experience. All IAs completed the evaluation within a single day.

6.3 *Image Sequence*

The order in which IAs view the imagery may subconsciously bias their responses. To address this, two different pseudorandom viewing orders were established for the imagery. Image order was random, with the exception of the original, uncompressed image, which was always the last one of that set seen by the IA. By forcing the uncompressed image to be the last one viewed, this pseudorandom order minimized the learning that an IA may unconsciously apply to the compressed renditions of a particular scene. The IAs did not know the image viewing order.

7 IA Selection

This study focused on the ability of military IAs to extract intelligence value from the compressed imagery. These analysts receive specialized training in image interpretation, target detection, and target identification for different classes of military targets or orders of battle, such as ground order of battle and electronic order of battle, and for different imaging sensors, such as SAR and electro-optical sensors. They are also trained to evaluate imagery very quickly. Although many military IAs may not have the formal, academic training of many contractor/civilian IAs, these military IAs are often highly skilled. However, their tenure as IAs may only be for a few months or for a few years, so they may not have the depth of experience that a civilian IA accrues during his or her career.

The IAs selected for this evaluation represented a cross section of the military IA population, with varying backgrounds and levels of experience. A total of 12 IAs participated in the evaluation. Four analysts were from the Enhanced Tactical Radar Correlator 525th Military Brigade, Fort Bragg, North Carolina. Four analysts were from the Air Force Contingency Airborne Reconnaissance System Deployable Ground Segment (CARS/DGS), Langley Air Force Base, Virginia; and four analysts trained at the National Ground Intelligence Center (NGIC), Washington Navy Yard, Washington, D.C. Each of the three groups of

Table 3 Image quality statistics, showing the average rating and standard error for each of the image metrics.

Average ± Standard Error		Radar Characteristic							
Image and Compression		NIIRS	Shadows	Contrast	Sidelobes	Speckle	Blurring and Motion	LOC confidence	Background Quality
RAW	1:1	3.5±0.1	3.6±0.2	4.0±0.2	4.7±0.2	4.2±0.2	4.3±0.2	5.1±0.2	4.1±0.1
JPEG	50:1	3.0±0.1	3.3±0.2	3.2±0.2	4.2±0.2	3.4±0.2	3.3±0.2	4.4±0.2	3.6±0.1
W/TCQ	50:1	3.6±0.1	3.8±0.2	4.0±0.1	5.0±0.2	4.6±0.2	4.3±0.2	5.0±0.2	4.1±0.1
IBC	50:1	3.2±0.1	3.2±0.2	3.6±0.1	4.5±0.2	3.9±0.1	3.9±0.2	4.6±0.2	3.9±0.1
W/TCQ	100:1	3.3±0.1	3.4±0.2	3.5±0.1	4.6±0.2	4.0±0.2	3.6±0.2	4.7±0.1	3.7±0.1
IBC	100:1	3.3±0.1	3.6±0.2	3.7±0.1	4.0±0.2	4.2±0.2	4.1±0.2	4.6±0.2	3.8±0.1
Average Difference±Standard Error									
JPEG	50:1	−0.6±0.1	−0.3±0.2	−0.7±0.2	−0.5±0.2	−0.9±0.2	−1.1±0.2	−0.7±0.1	−0.6±0.1
W/TCQ	50:1	0.1±0.1	0.2±0.2	0.0±0.1	0.2±0.2	0.3±0.2	0.0±0.2	−0.1±0.1	−0.1±0.1
IBC	50:1	−0.2±0.1	−0.3±0.2	−0.4±0.2	−0.1±0.2	−0.2±0.2	−0.4±0.2	−0.4±0.2	−0.1±0.1
W/TCQ	100:1	−0.2±0.1	−0.2±0.2	−0.4±0.2	−0.1±0.2	−0.2±0.2	−0.7±0.2	−0.4±0.2	−0.4±0.1
IBC	100:1	−0.2±0.1	0.1±0.2	−0.3±0.2	0.2±0.2	0.1±0.2	−0.2±0.2	−0.4±0.2	−0.1±0.1

Statistics are listed for uncompressed images, imagery compressed at 50-to-1, imagery compressed at 100-to-1. The average difference is the difference of each compression product from the uncompressed image's rating, with negative values indicating a degradation in image quality with respect to the uncompressed image.

IAs possessed comparable experience levels, ranging from 6 months to 6 years.

8 Analysis of Results

A variety of techniques were used to process[14] the responses of the IAs. Statistical screening of individual responses determined whether responses were atypical for each specific IA. Any outliers were excluded from subsequent analyses that examined the differences associated with the compression algorithms and compression ratios.

8.1 *Compression Effects on Image Quality*

A statistical analysis of the IA responses characterized the relative performance of the different compression algorithms. Table 3 shows the average and standard error statistics for the subjective measures (i.e., NIIRS, shadows, contrast, sidelobes, speckle, blurring and motion, LOC confidence, and background quality) as a function of image compression. The top portion of this table shows the simple averages. The standard error ranged from approximately 0.1 to 0.2. The lower portion of this table shows the average change in image rating, or Δ statistics, for each compression scheme, where negative values indicate a reduction in image quality. The Δ statistics were computed from the change in quality that each IA observed for the compressed renditions of each scene relative to the uncompressed version of the same scene.

Surprisingly, the uncompressed rendition was not always the highest performer. Instead, the 50-to-1 W/TCQ images rated slightly higher for Radar NIIRS, shadows, sidelobes, and speckle, although in most cases the difference was not statistically significant. The same was true for the few cases (shadows, sidelobes, and speckle), where 100-to-1 IBC rated higher than the uncompressed imagery. The slightly higher ratings for certain compressed products may reflect the fact that the compression processing results in some smoothing of the background and changes in speckle, which the IAs may find more visually pleasing. Further investigation is needed to fully understand this issue. Figure 4 illustrates the NIIRS and shadows results.

The JPEG imagery always rated the lowest by a significantly larger amount. The 50-to-1 W/TCQ imagery was overall the best performer of the compressed products, rating slightly higher than the 50-to-1 IBC imagery. Note, however, that the differences were statistically significant in all cases. An interesting result is that the 100-to-1 IBC often rated higher than the 50-to-1 IBC imagery, and also rated higher than the 100-to-1 W/TCQ in all but one case.

The quality of the JPEG imagery was lowest for blurring and motion artifacts (ΔBlur equal to -1.1), whereas most differences were -0.4 or less. The lower rating for JPEG imagery is probably due to the nature of JPEG compression, which creates a patchwork pattern of compression artifacts [Figure 2(d)].

Although IBC compression adds speckle to the compressed product, the speckle statistics for IBC imagery were very close to the statistics of the uncompressed imagery (within about ± 0.1). Table 3 shows that, on average, the speckle in 100-to-1 IBC is rated as higher quality than the uncompressed image, but this is not a statistically significant difference. In some instances (shadows, sidelobes, speckle) the ratings for IBC at 100-to-1 are higher than for IBC at 50-to-1 and the differences border on statistical significance. These findings appear counterintuitive, but might be a result of the greater smoothing implicit in the IBC 100-to-1 product. Further investigation will be required to test this hypothesis.

To summarize, the subjective measures of image quality and image interpretability indicate universally poor performance for the standard JPEG product (Table 3). At 50-to-1 compression, the IBC product performs slightly, but statistically significantly, below the uncompressed imagery for

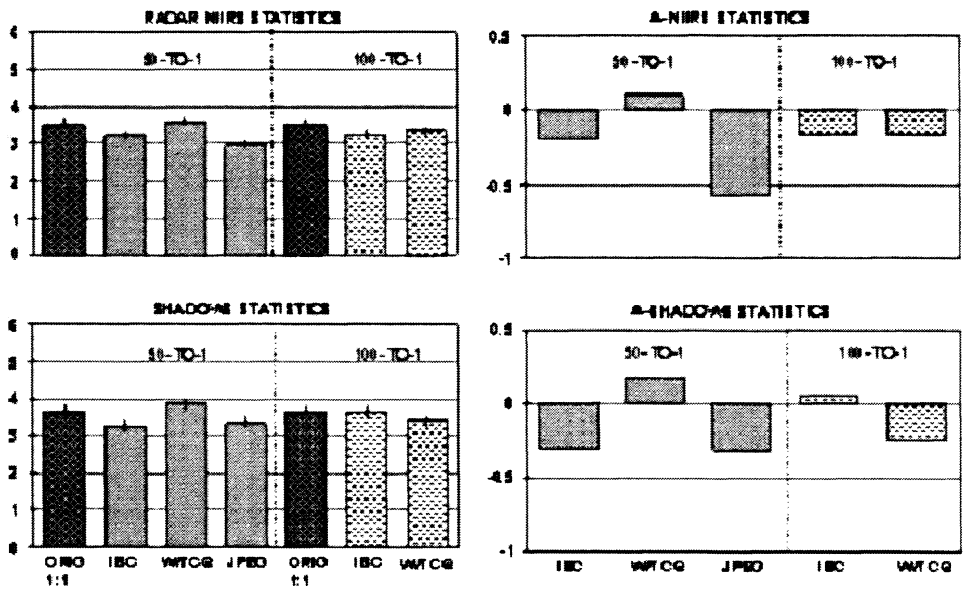

Fig. 4 Statistics of imagery analyst ratings of Radar NIIRS and shadows, for uncompressed images, imagery compressed at 50-to-1, and compressed at 100-to-1. Each column indicates the average rating for each image type. The vertical bar centered on the top of each column indicates ±1 standard error. The Δ statistics illustrate the average difference of each compression product from the uncompressed image's rating, with negative values indicating a degradation in image quality with respect to the uncompressed image.

certain image quality metrics. The W/TCQ 50-to-1 compression is statistically indistinguishable from the original imagery. For 100-to-1 compression, IBC and W/TCQ are comparable, with both exhibiting some slight loss in quality compared to the uncompressed original.

8.2 *Compression Effects on Image Exploitation*

The preceding section addressed the metrics that characterize image quality and image interpretability for the various compression products. Although these metrics can be reliable indicators of the potential utility of the imagery, they rest on the analysts' perceptions of features within the image. This section presents objective measures of image utility derived from the actual performance of the imagery analysts for specific image exploitation tasks. The IAs' responses were compared to image truth to derive the objective performance for

1. target detection, as measured by the ability of the IA to count military vehicles in a designated region of interest
2. target classification and identification for designated vehicles in the images
3. identification of common background features in designated regions of the images

Objective performance was analyzed similarly to image quality. First, the data were screened to flag possible outliers. In this case, only two observations for vehicle identification were set aside because the observed values exceeded the allowable range. All remaining data were used to assess the performance differences as a function of compression product, compression ratio, and IA, using a standard analysis of variance. The analysis of variance tested for differences due to compression while controlling for other factors, namely, IA and compression ratio.[15] When significant effects were found, *post hoc* testing indicated the statistical relationship among the compression products.

8.2.1 *Object count*

The ability to count military vehicles was scored by comparing the observed count to the image truth for each scene. IAs were penalized for both undercounting and overcounting. The performance metric, however, is slightly asymmetric, particularly for small numbers of targets. This is consistent with many military missions where undercounting of enemy targets could be more serious than overcounting. The object count performance metric M was

$$M=\left|\log\left[\frac{\text{observed}-0.001}{\text{truth}}\right]\right|.$$

The logarithm down weights extreme observations, to avoid undue influence in the analysis of variance. The addition of the small constant, while introducing a slight asymmetry, avoids taking a log of zero for the rare cases where reported vehicle counts were zero. A perfect score is 0 (zero).

The analysis revealed a significant difference only among the 50-to-1 compression products (Table 4). In this case, the ranking of products, from best to worst, is

1. original image
2. W/TCQ
3. IBC
4. JPEG

Table 4 Objective measures of image quality.

Image and Compression		Vehicle Measures: Count Accuracy	Vehicle Measures: Classification: Wheeled/Tracked	Vehicle Measures: ID by Type	Background Features: Accuracy	Background Features: Completeness
RAW	1:1	0.50	0.41	0.28	0.63	0.46
JPEG	50:1	1.32	0.23	0.09	0.64	0.48
W/TCQ	50:1	0.54	0.32	0.22	0.63	0.48
IBC	50:1	1.00	0.30	0.17	0.69	0.47
W/TCQ	100:1	0.60	0.32	0.24	0.66	0.47
IBC	100:1	0.87	0.35	0.21	0.64	0.45

Vehicles were used to measure count accuracy *M*, the proportion correctly classified (wheeled versus tracked), and the proportion correctly identified by type. The count accuracy is 0 for a perfect score and penalizes for both undercounting and overcounting. Background features are quantified using accuracy, i.e, the ratio of the number correct to the number reported, and by completeness, i.e., the ratio of the number correct to the total in the image.

The differences were statistically significant, and this is the only instance where IBC and W/TCQ differed significantly. For the 100-to-1 compression, the differences in performance, while suggesting a similar pattern, were not statistically significant. One possible reason for the difference with IBC could be the additional speckle in the IBC rendition, which may have confused the vehicle counting process.

8.2.2 *Vehicle classification and identification*

The vehicle classification and identification tasks were performed in two steps. The IA's attention was directed to a particular vehicle or group of vehicles in the image. The IA first was asked to determine if the vehicles were wheeled or tracked. Then the IA was asked to determine the type of vehicle; e.g., tank, armored personnel carrier (APC), truck, if possible. The IA's answer was scored against ground truth. The performance metrics were the proportion correctly classified (wheeled versus tracked) and the proportion correctly identified by type.

For both vehicle classification and identification, performance differed significantly among the three 50-to-1 compression products (Table 4). *Post hoc* testing revealed that the significant effect was due to the poor performance by JPEG. Performance differences among IBC, W/TCQ, and the original product were not statistically significant. Similarly, for the 100-to-1 compression, performance differences between the two compression products were not statistically significant.

8.2.3 *Context information*

The background areas of the evaluation imagery consisted of common features, such as rocks, dirt, low vegetation, forests, roads, and streams. The IAs were asked to indicate all such features found in the designated regions. The imagery analysts' responses were compared to the image truth to determine both the accuracy and the completeness of this exploitation task. The performance metrics were defined by

$$\text{accuracy} = \frac{\text{number correct}}{\text{total reported}},$$

and

$$\text{completeness} = \frac{\text{number correct}}{\text{truth total}}.$$

The results indicated that identification of background features is relatively unaffected by the compression algorithms (Table 4). The differences among the compression products are small and not statistically significant.

9 Summary and Conclusions

The image quality metrics (Radar NIIRS, shadows, speckle, etc.) show that some aspects of image quality are statistically significantly different from the uncompressed imagery (Table 5). In particular, the 50-to-1 JPEG algorithm consistently rated much lower for all metrics. Evaluation of the 50-to-1 JPEG product indicates that JPEG compression should not be used at higher compression ratios, and 100-to-1 JPEG was not evaluated in this study for this very reason.

Statistically significant differences among the compression products were observed only for the 50-to-1 compression for the vehicle detection and identification tasks (Table 6). In general, IAs performed slightly better on vehicle detection and recognition tasks using the original imagery. Performance using the IBC and W/TCQ products was generally similar, while performance on the JPEG product was universally worse. For identification of background features, performance was essentially constant across the full set of imagery products, including the original uncompressed imagery.

These results indicate that, with the exception of JPEG compression, even 100-to-1 compression had minimal impact on the intelligence value of the imagery as assessed by human IAs. Both IBC and W/TCQ compressions seem to preserve target information, and the nature of background degradation depends on the compression algorithm itself. These results indicate that image compression at even higher ratios will probably preserve intelligence value, at least for some military missions.

This study demonstrates that tactical SAR imagery can be highly compressed (up to 100-to-1) with negligible loss in imagery quality or utility. Furthermore, the concept of differential compression rates, as embodied in IBC, appears viable. These findings suggest that a hybrid compression

Table 5 Summary of results for subjective image quality ratings, indicating whether statistically significant results occurred for the different image features and for different compression algorithms and ratios.

Performance Metrics	50-to-1 Compression		100-to-1 Compression	
	Result	Source of Difference	Result	Source of Difference
Radar NIIRS	Significant	JPEG	Significant	IBC
Shadows	Significant	JPEG, IBC	Not significant	
Contrast	Significant	JPEG, IBC	Significant	W/TCQ
Sidelobes	Significant	JPEG	Not significant	
Speckle	Significant	JPEG, IBC	Not significant	
Blurring and motion	Significant	JPEG	Significant	W/TCQ
LOC confidence	Significant	JPEG, IBC	Significant	IBC, W/TCQ
Background quality	Significant	JPEG	Significant	IBC, W/TCQ

algorithm could be implemented that uses W/TCQ with differential compression rates for regions of interest versus background. In principle, such an algorithm should exhibit the strengths of both approaches. Future analysis should investigate the performance associated with such algorithms. Furthermore, future evaluations should explore the practical limits of SAR compression, to determine the compression rates at which substantial loss in intelligence information occurs.

10 Appendix: Radar National Imagery Interpretability Rating Scale[4]

Rating Level 0

Interpretability of the imagery is precluded by obscuration, degradation, or very poor resolution.

Rating Level 1

Detect the presence of aircraft dispersal parking areas.
Detect a large cleared swath in a densely wooded area.
Detect, based on presence of piers and warehouses, a port facility.
Detect lines of transportation (either road or rail), but do not distinguish between.

Rating Level 2

Detect the presence of large (e.g., BLACKJACK, CAMBER, COCK, 707, 747) bombers or transports.
Identify large phased array radars (e.g., HEN HOUSE, DOG HOUSE) by type.
Detect a military installation by building pattern and site configuration.
Detect road pattern, fence, and hardstand configuration at SSM launch sites (missile silos, launch control silos) within a known ICBM complex.
Detect large noncombatant ships (e.g., freighters or tankers) at a known port facility.
Identify athletic stadiums.

Table 6 Summary of results for objective performance metrics, indicating whether statistically significant results occurred for the different image features and for different compression algorithms and ratios.

Performance Metrics	50-to-1 Compression	100-to-1 Compression
Vehicles		
Count	Significant	Not significant
Classification	Significant	Not significant
Identification	Significant	Not significant
Background		
Accuracy	Not significant	Not significant
Completeness	Not significant	Not significant

Rating Level 3

Detect medium-sized aircraft (e.g., FENCER, FLANKER, CURL, COKE, F-15).
Identify an ORBITA site on the basis of a 12-m dish antenna normally mounted on a circular building.
Detect vehicle revetments at a ground forces facility.
Detect vehicle/pieces of equipment at a SAM, SSM, or ABM fixed missile site.
Determine the location of the superstructure (e.g., fore, amidships, aft) on a medium-sized freighter.
Identify a medium-sized (approximately six track) railroad classification yard.

Rating Level 4

Distinguish between large rotary-wing and medium fixed-wing aircraft (e.g., HALO helicopter versus CRUSTY transport).
Detect recent cable scars between facilities or command posts.
Detect individual vehicles in a row at a known motor pool.
Distinguish between open and closed sliding roof areas on a single bay garage at a mobile missile base.
Identify square bow shape of ROPUCHA class (LST).
Detect all rail/road bridges.

Rating Level 5

Count all medium helicopters (e.g., HIND, HIP, HAZE, HOUND, PUMA, WASP).
Detect deployed TWIN EAR antenna.
Distinguish between river crossing equipment and medium/heavy armored vehicles by size and shape (e.g., MTU-20 versus T-62 MBT).
Detect missile support equipment at an SS-25 RTP (e.g., TEL, MSV).

Distinguish bow shape and length/width differences of SSNs.
Detect the break between railcars (count railcars).

Rating Level 6
Distinguish between variable and fixed-wing fighter aircraft (e.g., FENCER versus FLANKER).
Distinguish between the BAR LOCK and SIDE NET antennas at a BAR LOCK/SIDE NET acquisition radar site.
Distinguish between small support vehicles (e.g., UAZ-69, UAZ-469) and tanks (e.g., T-72, T-80).
Distinguish between the raised helicopter deck on a KRESTA II (CG) and the helicopter deck with main deck on a KRESTA I (CG).
Identify a vessel by class when singly deployed (e.g., YANKEE I, DELTA I, KRIVAK II FFG).
Detect cargo on a railroad flatcar or in a gondola.

Rating Level 7
Identify small fighter aircraft by type (e.g., FISHBED, FITTER, FLOGGER).
Distinguish between electronics van trailers (without tractor) and van trucks in garrison.
Distinguish, by size and configuration, between a turreted, tracked APC and a medium tank (e.g., BMP-1/2 versus T-64).
Detect a missile on the launcher in an SA-2 launch revetment.
Distinguish between bow mounted missile system on KRIVAK I/II and bow mounted gun turret on KRIVAK III.
Detect road/street lamps in an urban, residential area or military complex.

Rating Level 8
Distinguish the fuselage difference between a HIND and a HIP helicopter.
Distinguish between the FAN SONG E missile control radar and the FAN SONG F based on the number of parabolic dish antennas (three versus one).
Identify the SA-6 transloader when other SA-6 equipment is present.
Distinguish the limber hole shape and configuration differences between DELTA I and YANKEE I (SSBNs).
Identify the dome/vent pattern on rail tank cars.

Rating Level 9
Detect major modifications to large aircraft (e.g., fairings, pods, winglets).
Identify the shape of antennas on EW/GCI/ACQ radars as parabolic, parabolic with clipped corners, or rectangular.
Identify, based on presence or absence of turret, size of gun tube, and chassis configuration, wheeled or tracked APCs by type (e.g., BTR-80, BMP-1/2, MT-LB, M113).
Identify the forward fins on an SA-3 missile.
Identify individual hatch covers of vertically launched SA-N-6 surface-to-air system.
Identify trucks as cab-over-engine or engine-in-front.

Acknowledgments

This project was sponsored by the Defense Advanced Research Projects Agency (DARPA) under Contract No. F33615-95-C-1643, as part of the MSTAR Program directed by Dr. Robert Hummel. The study was executed by Science Applications International Corporation (SAIC). The authors wish to acknowledge the technical assistance provided by Mr. Louis Jaeger, Mr. David Webb, and Mr. David Cannon and the support of the imagery analysts provided by the United States Army and United States Air Force. The views expressed in this paper are those of the authors and do not necessarily represent the position of DARPA or SAIC.

References

1. Y. Niho and R. Hummel, "Intelligent bandwidth compression for SAR imagery transmission," DARPA Technical Report (1998).
2. R. W. Ives, P. Eichel, and N. Magotra, "Application of pixel segmentation to the low rate compression of complex SAR imagery," presented at IEEE IGARSS 1998, July, Seattle, WA.
3. R. W. Ives, P. Eichel, and N. Magotra, "Intelligent low rate compression of speckled SAR imagery," presented at IEEE Radar 1998, May, Dallas, TX.
4. R. W. Ives, N. Magotra, and G. D. Manyam, "Lossless compression of synthetic aperture radar images," presented at IEEE ISCAS 1996, May, Atlanta, GA.
5. P. Eichel and R. W. Ives, "Very low rate compression of speckled SAR imagery," *Proc. SPIE* **3391**, 118–125 (1998).
6. *Image Compression Users Handbook*, National Imagery Transmission Format Standard, National Imagery and Mapping Agency, (1998).
7. J. W. Owens, M. W. Marcellin, B. R. Hunt, and M. Kleine, "Compression of synthetic aperture radar video phase history data using trellis coded quantization techniques," *IEEE Trans. Geosci. Remote Sens.* **37**(2), 1080–1085 (1999).
8. J. W. Owens and M. W. Marcellin, "Rate allocation for spotlight SAR phase history data compression," *IEEE Trans. Image Process.* **8**(11), 1527–1533 (1999).
9. T. R. Fischer, M. W. Marcellin, and M. Wang, "Trellis coded vector quantization," *IEEE Trans. Inf. Theory* **37**, 1551–1566 () (Nov. 1991).
10. J. M. Irvine, "National imagery interpretability rating scales (NIIRS): overview and methodology," *Proc. SPIE* **3128**, 93–103 (1997).
11. J. M. Irvine and J. C. Leachtenauer, "A methodology for developing image interpretability rating scales," in *Proc. Am. Soc. Photogram. Remote Sens.* (Apr. 1996).
12. J. C. Leachtenauer, "National imagery interpretability rating scales: overview and product description," in *Proc. Am. Soc. Photogram. Remote Sens. Annu.* M. (Apr. 1996).
13. L. M. Maver, C. D. Erdman, and K. Riehl, "Imagery interpretability ratings scales," *Society for Information Display Symp. Digest* **26**, 117–120 (May 1995).
14. B. J. Winer, *Statistical Principles in Experimental Design*, 2nd ed., McGraw-Hill, New York, (1971).
15. H. Sheffe, *The Analysis of Variance*, John Wiley & Sons, New York (1959).

John M. Irvine is the Director of Imagery Systems at Science Applications International Corporation (SAIC). He has served as chief scientist for several programs to evaluate assisted image exploitation systems, to develop and evaluate automatic target recognition (ATR) and image understanding technology, and to develop novel biometric techniques for human identification. Prior to joining SAIC, Dr. Irvine conducted numerous studies of image quality and image utility at the Environmental Research Institute of Michigan (ERIM). He has over 20 years experience in applied statistics, modeling, and test and evaluation, with an emphasis on applications of remote sensing and image processing. He holds a PhD in mathematical statistics from Yale University.

Barbara A. Eckstein is a senior systems consultant with McClendon Automation, Inc. She currently works at the National Imagery and Mapping Agency (NIMA) in Washington, D.C., concentrating on exploiting multi- and hyperspectral commercial imagery. Dr. Eckstein's expertise includes remote sensing, automatic target recognition, image compression, systems engineering and integration, and physical oceanography. She holds a PhD in geophysics from the University of Wisconsin–Madison as well as bachelor's and master's degrees in physics. Dr. Eckstein also interned at The National Aeronautics and Space Administration Johnson Space Center and at Amoco Production Company, and was a postgraduate researcher at Scripps Institution of Oceanography at the University of California–San Diego, studying mesoscale eddies in the California current system.

Robert A. Hummel is a program manager in the Information Exploitation Office of the Defense Advanced Research Projects Agency (DARPA), where he has directed a number of programs including intelligent bandwidth compression (IBC), the moving and stationary target acquisition and recognition (MSTAR), and JIGSAW. Dr. Hummel's research interests include automatic target recognition, computer vision, and image processing. He has served on the Naval Studies Board and consulted with the Defense Science Board on topics of target recognition. He holds a PhD in mathematics from the University of Minnesota and served on the faculty at New York University in the Computer Science Department from 1980 through 1995.

Richard J. Peters received his BS degree in electrical engineering and his MS degree in computer science from the George Washington University in 1986 and 1991, respectively. After developing and evaluating computer vision algorithms for the U.S. Army Night Vision Laboratory, he is currently an engineer for Science Applications International Corporation (SAIC) integrating hyperspectral sensor simulations. His technical interests include computer vision, spatially variant image compression, synthetic aperture radar, and hyperspectral imaging.

Rhonda Ritzel is a senior systems engineer with Science Applications International Corporation (SAIC) and has over 27 years experience in systems engineering and integration activities associated with the intelligence exploitation process, including tactical, theater, and national systems operations. Ritzel has been involved in several C4ISR programs and projects in support of the Department of Defense, and in particular, the Defense Advanced Projects Agency (DARPA). Ms. Ritzel is a retired U.S. Army military intelligence officer with extensive experience in the field of imagery intelligence.