Development of a Motion Imagery Quality Metric

John M. Irvine[1]; Charles Fenimore[2]; David Cannon[1]; Donna Haverkamp[1]; John Roberts[2]; Steven A. Israel[1]; Larry Simon[1]; James Miller[1]; Ana Ivelisse Avilés[2]; Michelle Brennan[3]

[1] SAIC; 20 Burlington Mall Road, Suite 130, Burlington, MA, 01803 (john.m.irvine@saic.com)
[2] National Institute of Standards and Technology; 100 Bureau Dr. MS 8940, Bldg 225 RM A265, Gaithersburg, MD 20899-8951
[3] Moriarty and Associates, 12300 Sunrise Valley Dr., Reston, VA 20191

ABSTRACT

The motion imagery community would benefit from the availability of standard measures for assessing image interpretability. The National Imagery Interpretability Rating Scale (NIIRS) has served as a community standard for still imagery, but no comparable scale exists for motion imagery. Several considerations unique to motion imagery indicate that the standard methodology employed in the past for NIIRS development may not be applicable or, at a minimum, require modifications. Traditional methods for NIIRS development rely on a close linkage between perceived image quality, as captured by specific image interpretation tasks, and the sensor parameters associated with image acquisition. The dynamic nature of motion imagery suggests that this type of linkage may not exist or may be modulated by other factors. A set of studies sponsored by the National Geospatial-Intelligence Agency (NGA) have been conducted to understand and quantify the effects of critical factors. Factors that could affect perceived interpretability of motion imagery that have been explored in these studies include:

- Target motion
- Camera motion
- Ground Sampled Distance (GSD), i.e., resolution
- Scene complexity
- Color
- Frame rate
- Image exploitation task category

This paper summarizes the findings from these evaluations. Based on the findings of these studies, we are formulating the plans for the development of a motion imagery quality metric.

Keywords: Image and video quality, image interpretability, motion imagery, NIIRS, full motion video

1. INTRODUCTION

The National Geospatial-Intelligence Agency (NGA) is conducting research and development into the feasibility of developing an interpretability scale for motion imagery. The National Imagery Interpretability Rating Scale (NIIRS) is a quantification of image interpretability that has been embraced by the Intelligence Community for still imagery (Irvine, 2003), (Leachtenauer, 1996), (Maver, *et,al,* 1995). Each NIIRS level indicates the types of exploitation tasks an image can support based on the expert judgments of experienced analysts. Development of a NIIRS for a specific imaging modality rests on a perception-based approach (Irvine, 1997). Accurate methods for predicting NIIRS from the sensor parameters and image acquisition conditions have been developed empirically and substantially increase the utility of NIIRS (Leachtenauer, *et al*, 1997). In exploring avenues for development of a similar metric for motion imagery, a clearer understanding of the factors that affect the perceived quality of motion imagery is needed. Several studies have explored specific isolated aspects of this problem, such as target motion, camera motion, color, and frame rate (Irvine, *et al*, 2004), (Irvine, *et al*, 2005), (Fenimore, *et al*, 2006), . This study examines the ability of imagery analysts to perform various exploitation tasks on motion imagery. Since a NIIRS-like scale is comprised of specific image exploitation tasks, understanding how these tasks relate to the imagery and image parameters is a vital step in developing a scale for motion imagery.

2. OBJECTIVES

This study consisted of two evaluations in which trained imagery analysts (IAs) rated their confidence in performing specific image exploitation tasks on a set of motion imagery clips. The tasks included tasks that could, in principle, be performed with still imagery, such as detection and recognition of various targets or objects. Other tasks included in the evaluation were specific to motion imagery, focusing on detection and recognition of activities. In the remainder of this paper, the term *static tasks* to refers to detection and recognition of objects and *dynamic tasks* refers to detection and recognition of activities. The specific objectives of the study were to explore the relationship between these tasks and properties of the motion imagery. Specific objectives were:

- Assess the relationship between satisfaction of the tasks and spatial resolution (GSD). Task "satisfaction" means that the exploitation task could performed with the motion imagery clip or one of comparable quality, if the relevant scene content were present. One interesting, related question is whether objects can be detected and identified at a lower GSD (compared to a still frame) because of motion-related cues.
- Explore the relationship between operator performance of these tasks and temporal resolution, as measured by frame rate. In particular, are both static and dynamic tasks sensitive to the frame rate? Studies measuring objective task performance with synthetic imagery show that performance degrades with decreasing frame rates (Fenimore, *et al*, 2006).
- Assess the satisfaction of motion-specific exploitation tasks for motion imagery. Historical experience in the development of NIIRS demonstrated that analysts could consistently assess performance of static tasks on still imagery. Can analysts consistently rate their satisfaction of tasks with motion imagery? We know from earlier studies that IAs can consistently rate motion imagery using the Visible NIIRS (Irvine, *et al*, 2004), (Irvine, *et al*, 2005). Operator consistent rating indicates that static tasks are likely to be rated consistently for motion imagery. Will the same holds true for the dynamic tasks?
- Explore the sensitivity of perceived criteria satisfaction to scene content. The motion imagery clips include cases with both high target motion and little or no target motion. An earlier evaluation suggested that there is a small, but significant perceptual quality difference due to target motion (Irvine, *et al*, 2004)]. Will this difference carry over to assessment of specific tasks?

3. DESIGN OF THE EVALUATIONS

Two evaluations provided the data to address the objectives described above. The first evaluation varied a number of factors to provide an initial assessment of each of the study objectives. The second evaluation held frame rate constant at 30 frames per second (fps) in order to develop a more statistically robust assessment of consistency and the relationship between task satisfaction and GSD.

3.1 Image Exploitation Tasks:

The starting point for developing the image exploitation tasks was the set of tasks or criteria that comprise the Visible NIIRS. These and other tasks used in previous NIIRS development efforts provided the static tasks. Development of dynamic tasks required identifying activities of interest on motion imagery. To relate image exploitation tasks to relevant missions for motion imagery, four scenarios of interest have been identified:

- Force protection
- Urban operations
- Targeting mobile and relocatable targets
- Situational awareness with respect to maneuver forces, i.e., intent

A list of approximately 50 dynamic tasks was compiled and each task was associated with the scenario or scenarios it addressed. These candidate tasks, both static and dynamic, were reviewed to select specific tasks for the evaluation that would span the range from easy to difficult. In addition, the tasks referenced common objects and activities that would be familiar to the analysts.

3.2 Imagery

To judge whether or not a NIIRS-like scale may be able to describe motion imagery and to assess the factors affecting perceived motion image quality, we must use image clips that cover a range of GSDs and exhibit both low and high motion. To explore the interaction of these factors with frame rate, it is necessary to also cover a range of frame rates. Frame rate variations were performed by frame sampling on the selected clips. A large set of imagery (approximately 500 image clips) were reviewed with respect to motion content, complexity, and quality. The imagery selected for the evaluation ranged in GSD from 200 inches to approximately 1 inch. We extracted 15-second segments for the evaluation. Each of these primary clips was then be subjected to frame rate sampling to produce 15-second renditions at 30, 15, and 1 fps. A single frame was selected from the middle of each 15-second clip to produce a still image for comparison, as well.

3.3 Evaluation 1:

The evaluation consisted of the assessment of each clip, at each of the frame rates and for a single frame, according to each task. For every clip at each frame rate, each analyst rated how effectively he/she believed each exploitation task could be performed. The ratings were on a scale of 0 (no confidence) to 100 (very high confidence). The analyst viewed the clip as many times as needed. The analyst would then review the first task and make an assessment relative to that clip. After rating the first task, the analyst would go on to the second task and so on. Once all of the tasks were rated with respect to the first clip, the analyst would bring up the second clip and repeat the process, continuing until all tasks and been rated relative to all clips. The order of presentation of the clips was grouped into blocks by frame rate and randomized within each block. The order of the frame rates was counterbalanced across analysts and the order of tasks within each clip was randomized.

In addition, each full-frame (30 fps) clip and each still image (single frame) was rated with the Visible NIIRS. This allowed us to assess any correlation between NIIRS- rating by the analyst and the confidence assigned to the tasks. Because both the clip and the still frame were NIIRS rated, we could examine the effects of motion on NIIRS. Thus, the evaluation itself consists of three distinct steps:

1. Assessment of all clips at each frame rate and still frame according to each criterion.
2. NIIRS ratings of the 30 fps clips.
3. NIIRS ratings of the still frames.

Because each task is rated relative to each clip, the number of ratings grows quickly with the number of tasks and clips. To keep the size of the evaluation manageable and to avoid fatigue, we constrained the evaluation to a relatively small number of clips and tasks. Six parent clips were used to generate renditions at 30 fps, 15fps, 1 fps, and the still image. Thus, 24 clips (6 parent clips times 4 frame rates) were presented to the analysts. Three of the parent clips had high target motion and three had little or no target motion (Table 1). The spatial resolutions included coarse, medium, and fine GSDs. Fourteen exploitation tasks were used: 7 static tasks and 7 dynamic ones.

Table 1. Imagery Data for Evaluation 1

	Resolution		
	Coarse (60-100” GSD)	**Medium** (10-60” GSD)	**Fine** (1-10” GSD)
Low Target Motion	30, 15, 1, and 0 fps	30, 15, 1, and 0 fps	30, 15, 1, and 0 fps
High Target Motion	30, 15, 1, and 0 fps	30, 15, 1, and 0 fps	30, 15, 1, and 0 fps

3.4 Evaluation 2:

Evaluation 2 followed a very similar structure, but did not explore variations in frame rate. The goal for the second evaluation was to expand the number of criteria and the number of clips to provide a larger sample for assessments of rater consistency. The larger pool of data also supported more extensive comparisons of task satisfaction to GSD.

This evaluation included 20 video clips, all viewed at 30 fps, and 20 exploitation tasks. Eleven analysts participated in the evaluation.

4. ANALYSIS AND RESULTS

4.1 Evaluation 1

The primary measure of performance was the confidence assigned to each task relative to each clip. The first evaluation permitted an initial investigation of the effects of frame rate, scene content (high vs. low target motion), and the nature of the task (static vs. dynamic). The Visible NIIRS ratings provided additional information about perceptions of the motion imagery.

The results show that task confidence does vary with frame rate (Figure 1). When the tasks are broken out by static and dynamic, the expected pattern emerges. Static tasks, which could in principle be performed with still imagery of sufficient quality, are insensitive to frame rate. The dynamic tasks, however, are sensitive to frame rate and confidence drops dramatically with frame rate (Figure 2). The pattern is not sensitive to the scene content in the sense as defined by target motion. Task confidence exhibits the same pattern for both high-target-motion and low-target-motion clips (Figure 3). This is a promising result in terms of scale development, because it suggests that analysts can assess information potential from a clip independent of the level of target motion. The pattern is also consistent across the three GSD bins represented in the data (Figure 4).

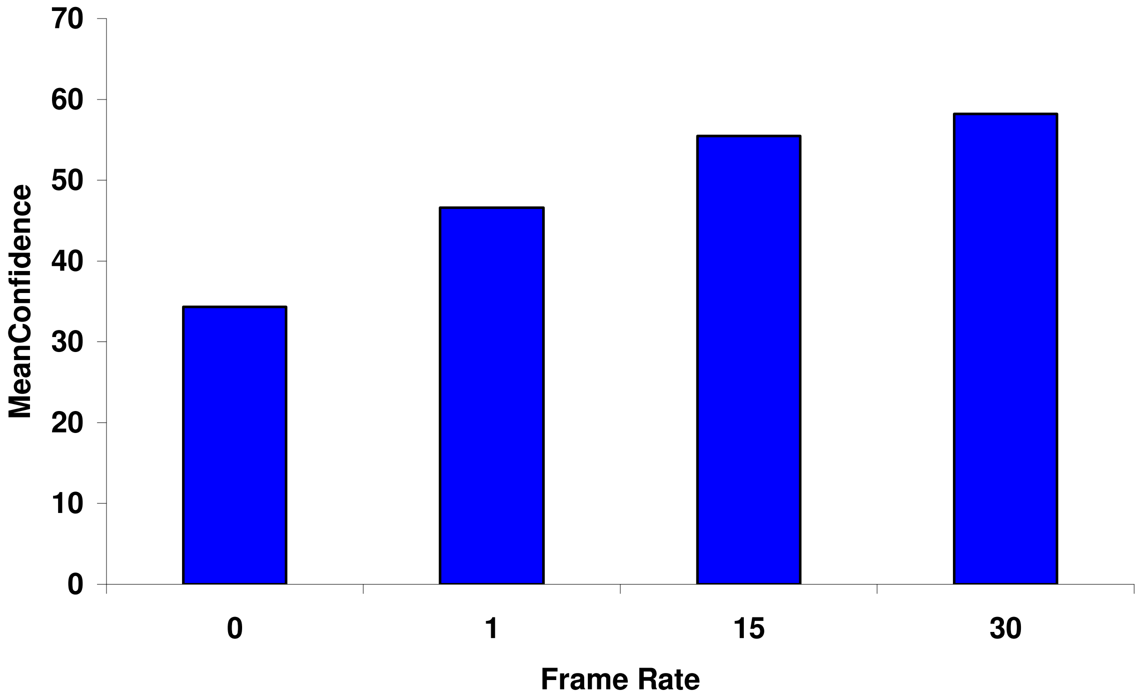

Figure 1. Mean Confidence Ratings Across All Tasks and Clip, By Frame Rate

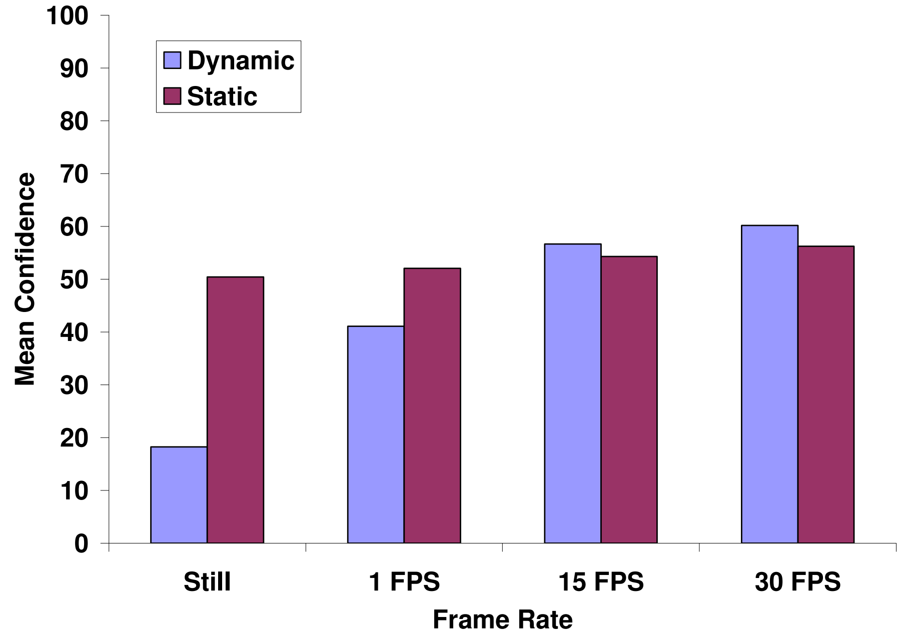

Figure 2. Mean Confidence Ratings Across All Tasks and Clip, By Frame Rate and Task Type

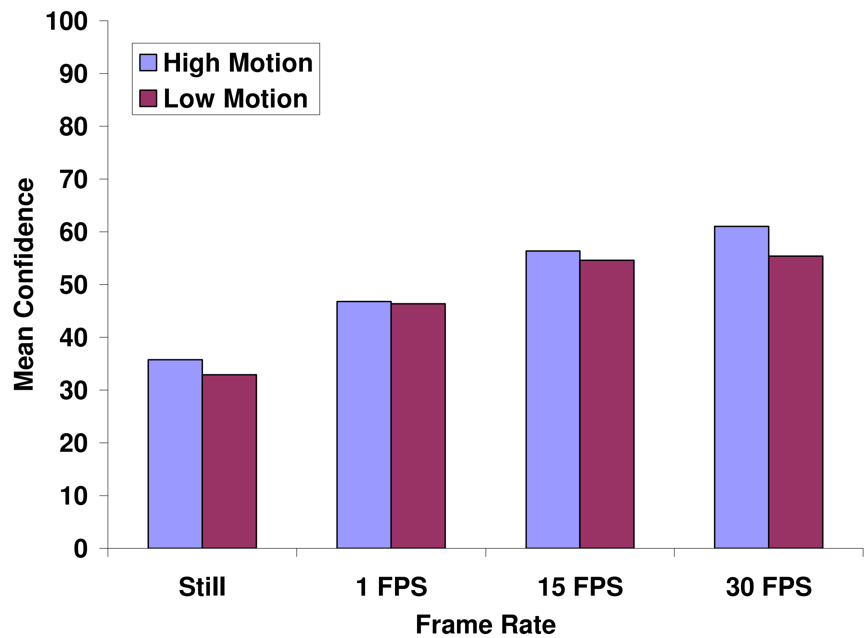

Figure 3. Mean Confidence Ratings Across All Tasks and Clip, By Frame Rate and Target Motion

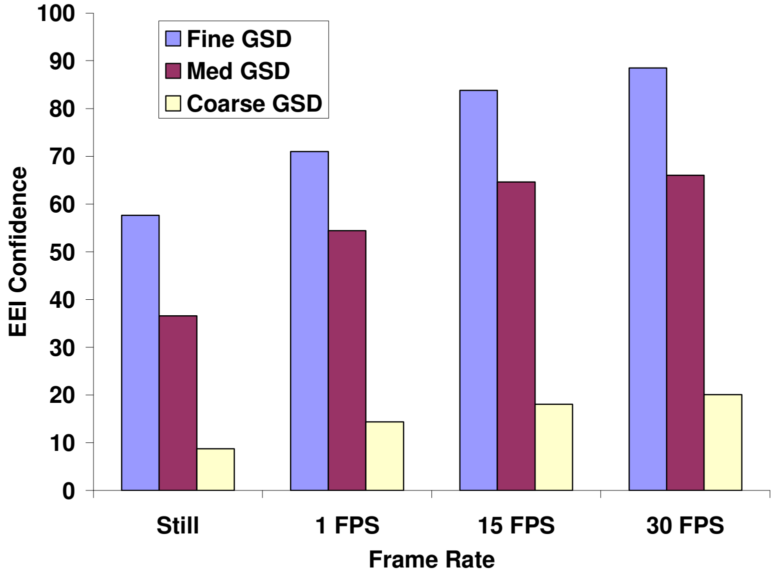

Figure 4. Mean Confidence Ratings Across All Tasks and Clip, By Frame Rate and GSD

Analysis of the Visible NIIRS ratings revealed the expected results. Visible NIIRS varies inversely with Log_{10}(GSD), as is also the case with still imagery (Figure 5). The pattern is independent of the level of target motion in the clip (Figure 5). It is also independent of frame rate (Figure 6). Both of these findings should be expected, since the Visible NIIRS addresses exploitation tasks for still imagery.

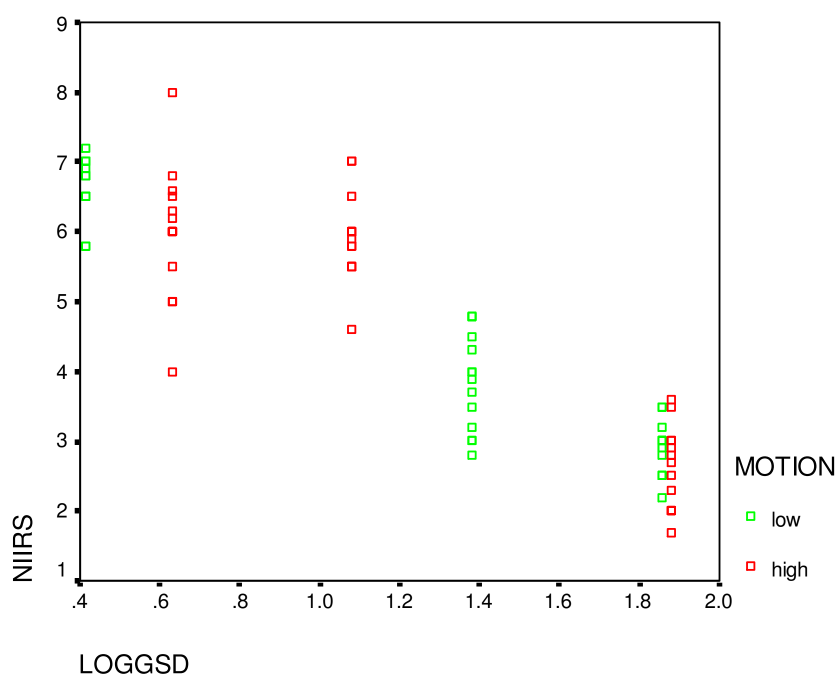

Figure 5. NIIRS Ratings by Log_{10}(GSD) and Target Motion (High vs. Low)

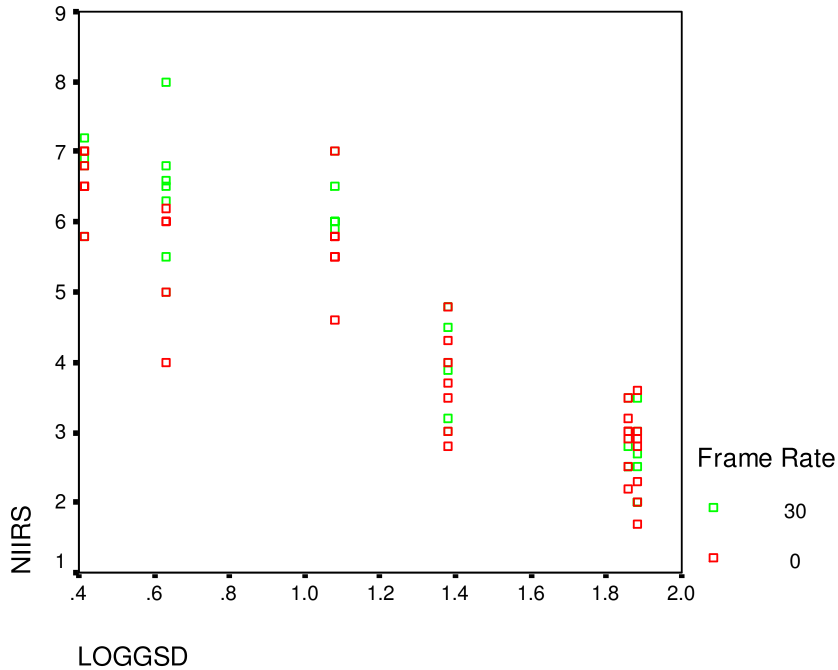

Figure 6. NIIRS Ratings by Log_{10} (GSD) and Frame Rate (30 fps vs. Still)

4.2 Evaluation 2

Again, the primary measure of performance was the confidence assigned to each task relative to each clip. The second evaluation provided a closer look at rater consistency and the relationship between task confidence and the properties of the image clip. In general, the analysts showed good agreement in their confidence ratings (Figure 7). Correlations between an individual's ratings and the mean of the group were roughly 0.9. In addition, the confidence ratings exhibit the desired relationship with GSD. As Figure 8 and 9 suggest, an "easy" task can be performed on most clips regardless of GSD.

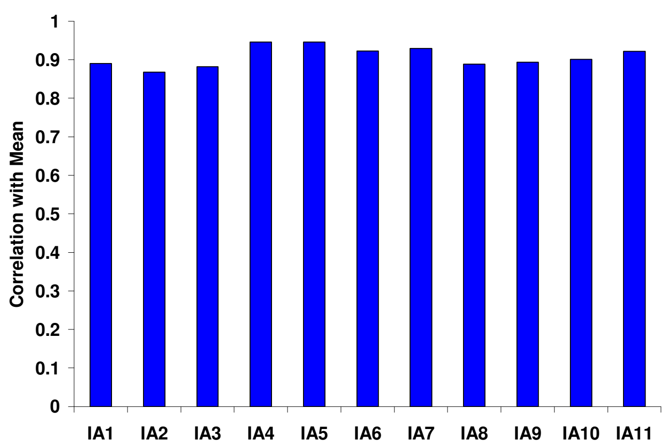

Figure 7. Task Confidence Ratings for Each Analyst: Correlation with the Mean

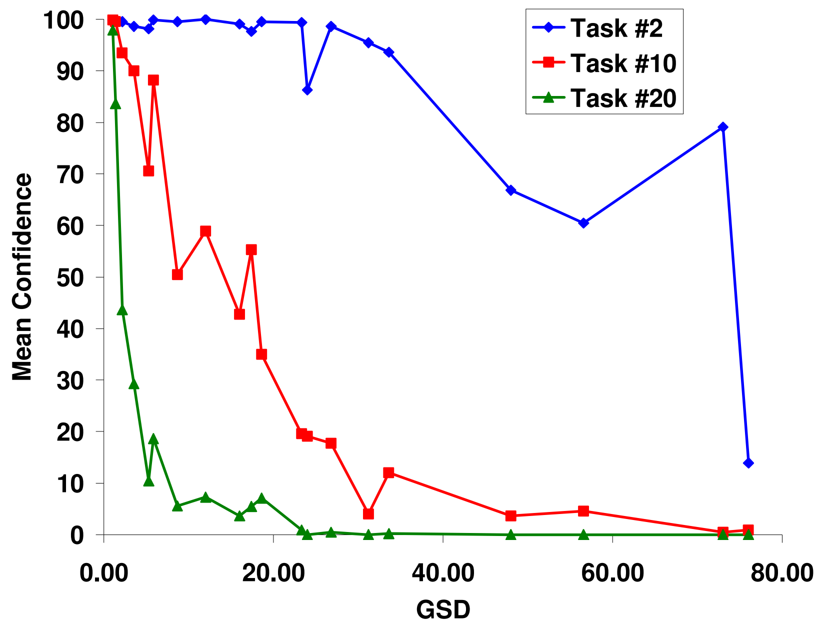

Figure 8. Task Confidence Ratings for Three (3) Tasks Relative to GSD for the Image Clips

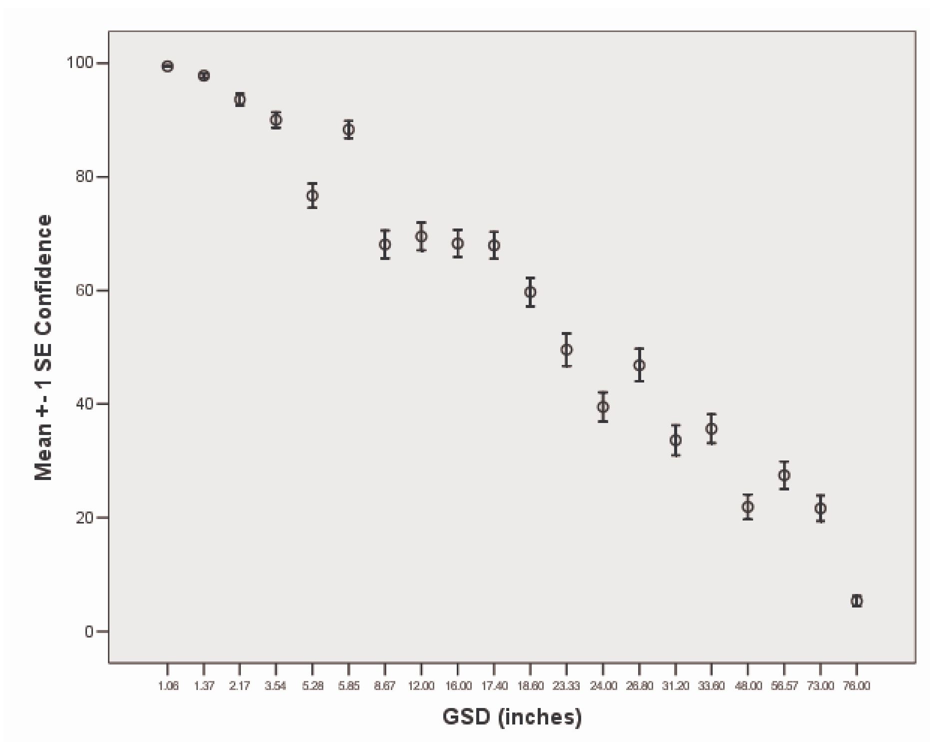

Figure 9. Task Confidence Ratings Across Tasks Relative to GSD for the Image Clips

5. CONCLUSIONS

These two evaluations provide valuable insight into analysts' perception of the interpretability of motion imagery. Following the approach used in the development of NIIRS, the interpretability can be defined by the types of image exploitation tasks that can be performed on a given clip. The findings from these two evaluations are consistent with this approach and point to three major conclusions:

1. The perceived ability to perform exploitation tasks depends primarily on the spatial resolution of the imagery, as measured by GSD
2. The ability to perform dynamic tasks, which involve detection and recognition of activities, depend on the temporal resolution, as measured by frame rate. Static tasks, however, are not sensitive to frame rate
3. Confidence in performance exploitation tasks, whether static or dynamic, does not depend on the level of target motion in the scene

These findings are consistent with the premise that development of a task-based scale for motion imager is feasible. Based on these findings, we have conducted a pilot investigation of the scale development methodology before embarking on the full scale development effort (Irvine, *et al*, 2006). Based on the experience to date, we conclude that development of a NIIRS-like scale for motion imagery is feasible and we hope to proceed with full scale development in the near future.

6. REFERENCES

J.M. Irvine,. (1997)"National Imagery Interpretability Rating Scales (NIIRS): Overview and Methodology" *Proceedings of the International Society for Optical Engineering (SPIE)*, 29-30 July, 1997, Volume 3128, pp.93-103.

J.M. Irvine (2003) "National Imagery Intelligence Rating Scale (NIIRS)" in Driggers, R.G. *The Encyclopedia of Optical Engineering*, Marcel Dekker., October 2003.

J.C. Leachtenauer (1996) "National Imagery Interpretability Rating Scales: Overview and Product Description," *Proceedings of the American Society of Photogrammetry and Remote Sensing Annual Meetings*, April 1996.

J. C. Leachtenauer, W. Malila, J. M. Irvine, L. Colburn, and N. Salvaggio, (1997) "General Image-Quality Equation: GIQE," *Applied Optics*, vol. 36, pp. 8322-8328, 1997.

L.A. Maver, C.D. Erdman, K. Riehl, (1995) "Imagery Interpretability Rating Scales", *Digest of Technical Papers, International Symposium Society for Information Display*, Society for Information Display, Vol. XXVI, pp: 117-120, 23 May. 1995.

John M. Irvine, Charles Fenimore, David Cannon, John Roberts, Steven A. Israel, Larry Simon, Charles Watts, James D. Miller, Ana Ivelisse Avilés, Paul F. Tighe, Richard J. Behrens, Michelle Brennan, Donna S. Haverkamp, (2005) "Factors Affecting Development of a Motion Imagery Quality Metric", *SPIE Defense and Security Symposium*, Orlando, FL, 28 Mar-1 Apr 2005.

John M. Irvine, Charles Fenimore, David Cannon, John Roberts, Steven A. Israel, Larry Simon, Charles Watts, James D. Miller, Ana Ivelisse Avilés, Paul F. Tighe, Richard J. Behrens (2004) "Feasibility Study for the Development of a Motion Imagery Quality Metric" *33rd Applied Imagery and Pattern Recognition Workshop: Image and Data Fusion,* IEEE Computer Society*,* Washington, 13-15 October 2004.

Charles Fenimore, John Irvine, David Cannon, John Roberts, Ivelisse Aviles, Steven Israel, Michelle, Brennan, Larry Simon, James Miller, Donna Haverkamp, Paul F. Tighe, Michael Gross (2006) "Perceptual Study of the Impact of Varying Frame Rate on Motion Imagery Interpretability"**,** SPIE Conference on Human Vision and Electronic Imaging XI, SPIE 6057-17, January 2006.

John M Irvine, David Cannon, James Miller, Jeffrey Bartolucci, Laurie Gibson, Charles Fenimore, John Roberts, Ivelisse Aviles, Michelle Brennan, Aloise Bozell, Larry Simon, Steven A. Israel (2006) "Methodology Study for Development of a Motion Imagery Quality Metric" SPIE Defense and Security Symposium, 17-21 April 2006, Orlando, Florida.