Validation of contrast and phenomenology in the Digital Imaging and Remote Sensing (DIRS) lab's Image Generation (DIRSIG) model

John E. Mason*, John R. Schott, Carl Salvaggio**, Joseph D. Sirianni***

Rochester Institute of Technology, Center for Imaging Science
Digital Imaging and Remote Sensing Laboratory
Chester F. Carlson Building, 54 Lomb Memorial Drive
Rochester, New York 14623-5604

*Currently with Eastman Kodak Company
**Currently with Hughes Aircraft
***Currently with E-Systems

ABSTRACT

Comparison of the components and the overall fidelity of infrared synthetic image generation models with truth data and imagery is a crucial part of determining model validity and identifying areas in which improvements can be made. The Rochester Institute of Technology's Digital Imaging and Remote Sensing Image Generation Model, DIRSIG, was validated in the midwave infrared (MWIR) and longwave infrared (LWIR) regions using measured meteorological, material, and radiometric data. Error propagation techniques clearly defined areas where improvements to the model could be made (e.g. inclusion of clouds). An overall comparison of truth and synthetic images yields RMS errors of as low as 1.8°C for actual temperature, and 5°C (LWIR) and 6°C (MWIR) for apparent temperatures. Analysis of rank order correlation statistic shows a very high correlation between brightness rank for object in the truth and DIRSIG images for most times of day.

1. INTRODUCTION

The ready availability of high speed computing has made the use of synthetic imagery attractive for a variety of applications. These include real-time flight simulation, algorithm development, training, mission rehearsal, and as an analyst aid. The model requirements vary considerably depending on the application. At one extreme there is a need in flight simulation for generation of images at very high rates of speed with reasonable geometric fidelity, but radiometric fidelity can usually be sacrificed. At the other extreme, for algorithm development or when used as an analyst's aid, we require very high radiometric fidelity but there is no need for real-time performance. The DIRSIG code was developed in response to this second need for models that would closely mimic reality. It is designed to produce synthetic images representing the radiance reaching an imaging sensor in any of up to 255 spectral bands located between 0.28 and 20.0 μm. The model is designed for high radiometric fidelity and attempts to model most significant phenomena using first principles physical models wherever possible. The DIRSIG model is described by Schott et al. 1992[1] and will only be briefly reviewed in the next section. This paper emphasizes efforts to validate synthetic image generation (SIG) models in the thermal infrared spectral region. Validation efforts aimed at assessing the mean level radiometric performance (i.e. RMS error between truth and SIG imagery) of DIRSIG are discussed at length by Mason et al. 1994[2] and Schott et al. 1993.[3] Our interest here will be focused on consideration of a metric aimed at evaluating the relative contrast produced in a synthetic image as compared to truth. The relative contrast, between a target and background or within a target, is a critical parameter in most image analysis as it is extensively used in both visual and machine-based interpretation. This paper describes how rank order correlation statistics can be used to evaluate the performance of a SIG model by comparing contrast levels in real and synthetic images. In addition, as part of the DIRSIG validation a number of phenomena were also identified as important based on their visual manifestation in the truth imagery. A number of these phenomena will also be briefly assessed in a non-quantitative manner in this paper to determine if the SIG model mimics the first order visual manifestation of these phenomena.

2. DIRSIG MODEL OVERVIEW

DIRSIG was originally written for the infrared region of the electromagnetic spectrum, but has since been expanded to include visible radiation. Its full range is 0.28-14 μm, and follows the image chain based primarily on first-order physical principles. A complete description of the model can be found in Schott et al. 1992.[1] The validation of DIRSIG reported here only covers the LWIR (8-13.3 μm) and MWIR (3-5 μm) portions of the spectrum.

0-8194-1593-6/94/$6.00

Fig. 1 gives a generalized block diagram of the submodels within DIRSIG. The geometry submodel contains the physical information of the objects in the target scene including spatial information (size, location, and orientation) and material attributes (e.g. emissivity, conductivity, etc.). The ray-tracer submodel relates the target scene to the sensor and the source (e.g. look-angle, time-of-day, and spectral range). Rays are cast from sensor to target to determine the type of interaction involved (e.g. specular bounce to sky, specular bounce to background, diffuse hit, or ray cast to sky) and to gather the pertinent parameters for computing object temperature and impinging radiance. The sensor submodel takes the spectral radiance information received at the front end of the sensor and passes it through a sensor spectral response function.

In the thermal submodel, surface temperatures of target objects are computed from temporal, meteorological, and material information. The thermal model incorporates a modified version of the Air Force Infrared Simulated Image Model (AIRSIM) known as THERM.[4,5] In the radiometry submodel, atmospheric contributions to the radiance impinging on the target and sensor (e.g. transmission, upwelled and downwelled radiance, etc.) are extracted from a radiometric database created particularly for the time, location, and atmospheric condition of the synthetic scene. The radiometric database comprises spectral vectors of radiance components extracted from the output of the Air Force Geophysics Laboratory (AFGL)/ Spectral Sciences Inc. MODTRAN Atmospheric Transmission/Radiance Code.[6,7]. The composite set of sub models is collectively capable of producing synthetic images that incorporate a wide range of the phenomena that impact the image formation process.

3. VALIDATION PROCEDURES

This section will describe the data collection and analysis procedures used on this effort, including a summary of the mean level temperature and radiometry error analysis.

3.1 Data Collection

Truth temperatures and imagery was collected in two experiments including all of the parameters necessary for input to THERM and DIRSIG. The experiments ran for forty-eight hour periods on the dates of October 5, 6, and 7, 1990 and June 22 and 23, 1992. A diagram of the June 1992 experiment is shown in Fig. 2. Imagery was captured from the roof of the RIT Center for Imaging Science building in order to simulate airborne sensor collections. LWIR imagery was collected using an Inframetrics IR camera containing a single Mercury/Cadmium/Telluride (HgCdTe) detector which was cooled by liquid nitrogen to a temperature of 77K. The image was scanned using electromechanical servos. During each time interval, five frames of imagery were grabbed using Werner Frei software on an Image Technology board. These frames were then averaged to reduce noise artifacts which appear in individual images.

MWIR imagery was also gathered during the June 1992 collection with a Platinum Silicide (PtSi) 2-D array video compatible IR imager supplied by the ARMY Night Vision Laboratory[8]. Visible images were also collected using a CCD camera for purposes of determining facet sun/shadow histories during the October 1990 collection. Sun/shadow history is used to better estimate object temperatures. During the June 1992 collection, four separate images in different bands (red, green, blue, and near infrared) were collected by the CCD camera using different filters. These images were used for a validation study of DIRSIG's predictions in the visible region of the electromagnetic spectrum. All images were collected every half hour, stored on computer hard drive, and backed up on magnetic tape.

Temperatures of objects in the scene were recorded by thermistors on fifteen minute intervals. The thermistors were calibrated using a water bath of known temperature. For both collections, object dimensions, angles and locations within the scene were measured, and radiosonde data was obtained from the Buffalo Airport for October 4-7, 1990 and June 22-24, 1992. To test THERM's weather parameter predictions, complete diurnal weather data from six days: June 23 and 24, 1987; October 6 and 7, 1987; and October 5 and 6, 1990, were used. Meteorological data were collected at fifteen minute intervals. The 1987 data were from an USAF collection at Wright Labs where similar ground truth data were available.

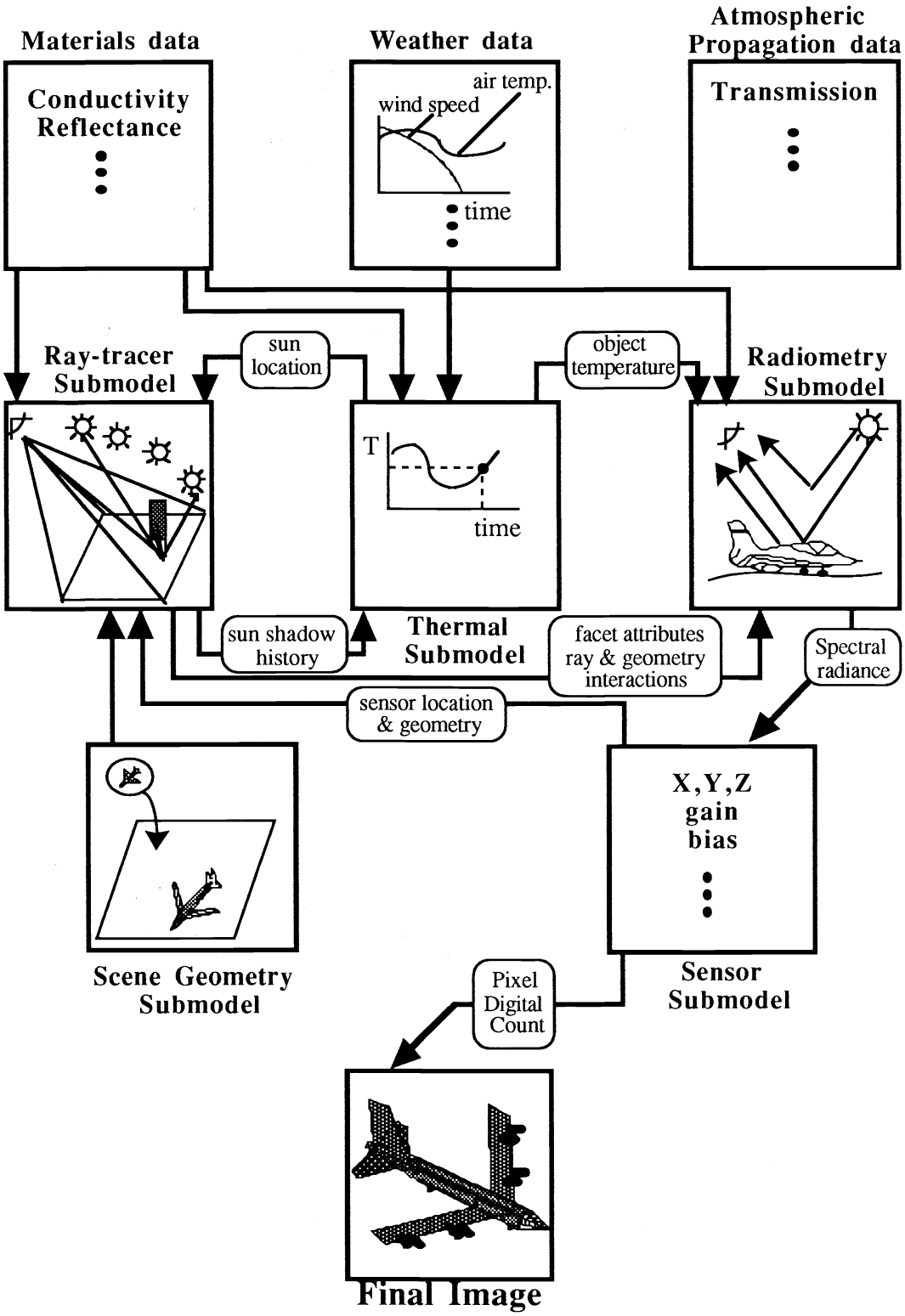

Fig. 1 Illustration of interactions between submodels in the DIRSIG model.

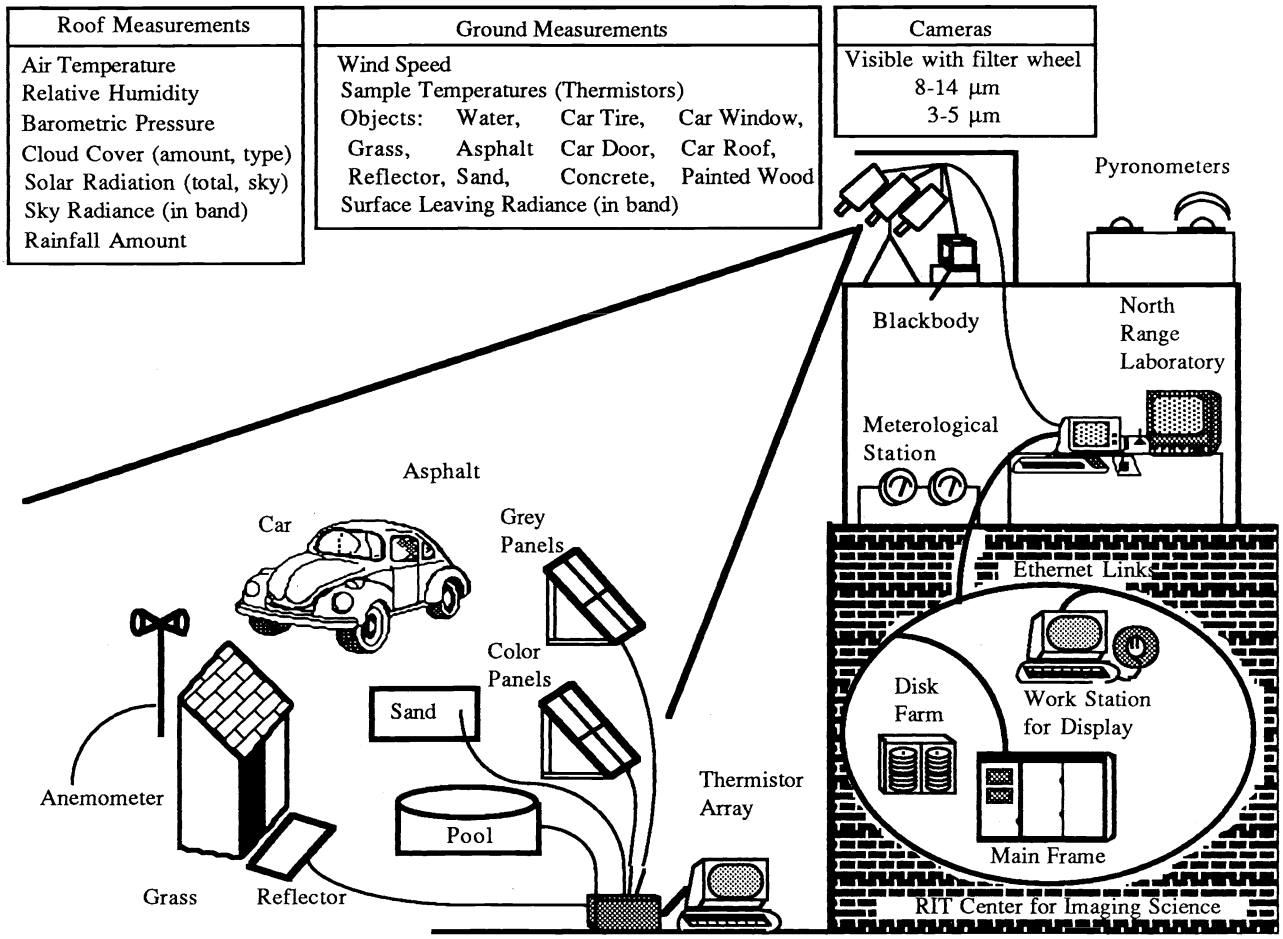

Fig. 2 Validation experiment.

3.2 Data Production

A geometric scene was constructed of the features imaged during the validation experiments. Synthetic images were produced representing the scene hourly over a twenty-four hour period. For targets in the scene with thermistors attached, the temperatures of the targets were also predicted by the SIG model for direct evaluation. The synthetic images were scaled through the same radiometric calibration parameters as the actual data so that radiance values or digital counts could be directly compared.

3.3 Data Analysis for Root Mean Square (RMS) Error

A detailed error propagation and radiometric error analysis using root mean square error assessment was conducted as reported by Mason et al. 1994[2]. Only a brief summary will be reported here. The error analysis of the temperature data showed an RMS error between truth and model of better than 2°C when the meteorological and material properties are well known and better than 4°C if the meteorological data are based on forecast data or the material parameters are generic textbook values. These data are the result of 13 samples compared every 15 minutes over 24 hours from data acquired on October 6, 1990 or October 6, 1990 and June 23, 1992. The sensitivity analysis indicated that the largest sources of temperature error due to the measured meteorological parameters were associated with lack of knowledge of wind speed and air temperature. However, when forecast meteorological data were used, the temperature error was dominated by lack of knowledge of solar and diffuse insolation levels driven by the inability to predict the presence or extent of cloud cover. The most important sources of temperature error due to object parameters were lack of knowledge of solar absorptivity, azimuthal orientation, and exposed area (i.e., the

fraction of the target in radiational exchange with the sky).

The RMS errors in radiance were computed in the MWIR and LWIR bands by comparing the truth and synthetic images of ten targets in hourly images over a 24 hour period. The RMS radiance errors expressed as apparent temperature were 5°C in the LWIR and 6°C in the MWIR. An error analysis showed that the largest sources of error were typically due to lack of knowledge of the atmospheric propagation parameters, the target emissivity, the temperature of the target, and the exposed area. Perhaps most importantly the analysis showed the need to properly account for the presence and location of clouds in both the temperature and radiance calculations. The radiance error analysis was performed on a day with a substantial cloud deck in the western sky for much of the day. The thermal model with measured data as the input can use this data to some extent, but it's treatment is currently location independent so east and west facing surfaces are treated the same. The radiometry model run in the validation did not include clouds as background objects and so this adds another substantial source of error.

An assessment of the RMS errors in DIRSIG indicates that the errors are surprisingly small for a first test of this type of model. The main purpose of this test was to identify limitations in the model in order to point to where future improvements would be productive. Thus, future improvements are focusing on inclusions of clouds, better treatment of exposed area, and a more complete treatment of directional emissivity (reflectivity). These efforts are targeted at improvements aimed at a substantial reduction in RMS errors based on a better understanding and treatment of the scene physics.

3.4 Data Analysis for Rank Order Correlation

In parallel with the RMS error analysis, an analysis of the SIG process was conducted aimed at assessing the model's fidelity for producing scenes appropriate for visual assessment and/or contrast based machine assessment. For these types of analysis, the mean level error (expressed as RMS error) may be largely irrelevant. Rather, the relative contrast of each object with the next is the important fidelity parameter. This is perhaps best seen in the case of an overall scene gain or bias change. If a synthetic scene perfectly represents an actual scene, it would have zero RMS error. If the model introduced a constant bias error of 5°C when expressed as apparent temperature, the RMS error would increase to 5°C. However, an analyst or a contrast based algorithm would consider either image equally acceptable since the relative contrast rank of each object remains essentially the same (i.e., each object's brightness ranking remains the same). From this assessment, we see that model errors that affect image-wide gain and bias levels will have little or no impact on image fidelity for contrast based analysis. However, gain and bias errors will produce substantial RMS errors that measure the limitations of SIG for absolute radiometric predictions.

In order to evaluate the relative brightness of the DIRSIG versus the truth images, 22 objects in the scene were selected for comparison. In each scene the objects were ranked in terms of brightness. Any difference in ranking between the DIRSIG image and the truth image is indicative of a contrast reversal for that object in the DIRSIG scene. The difference in the rank order value between an object in the truth image and the rank of the same object in the DIRSIG image is indicative of the number of contrast reversals that will exist for that object.

This concept can be viewed graphically by plotting the rank order in the truth scene versus the rank order in the DIRSIG scene for each object in the test. This is shown in Figure 3 for the LWIR image taken at 1200 hours. The corresponding truth and DIRSIG images are shown in Fig. 4. Unlike in the reflective regions of the electro-magnetic spectrum where this ranking remains fairly constant over time (except when specular glints occur), the rankings change continuously throughout the day in the thermal regions. An object that appears cold at one time of day (TOD) will be hot at another TOD and in the mid levels at a third TOD. Thus, a test of the rank order relationship of many objects in a scene will be a strenuous test of the performance of the SIG model.

The Spearman rank order correlation is used to measure the rank order relationships between two treatments[9]. In our case the truth scene and DIRSIG are compared. The rank order correlation is defined as

$$\rho_{TOD} = 1 - \frac{6\Sigma(R_i - R_i')^2}{n^3 - n} \tag{1}$$

where ρ_{TOD} is the correlation coefficient for each image pair at the (TOD) in the subscript, n is the number of samples, R_i is the rank in the truth image, and R_i' is the rank in the DIRSIG image for the i^{th} object. The

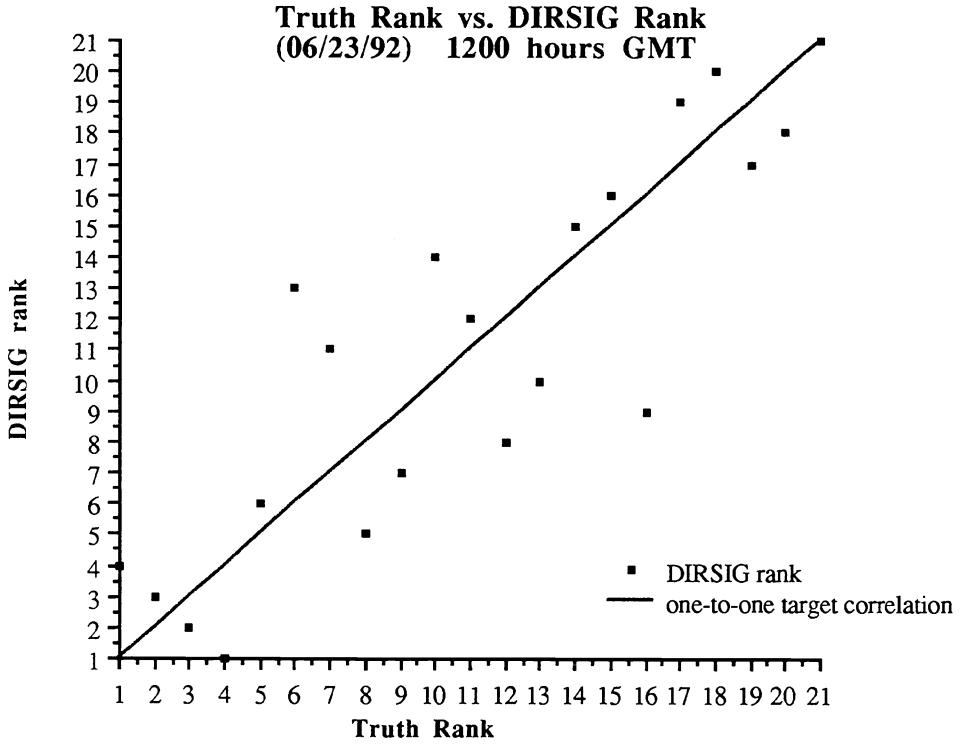

Fig 3. Plot of truth rank order versus DIRSIG rank order for 21 targets in a LWIR validation scene.

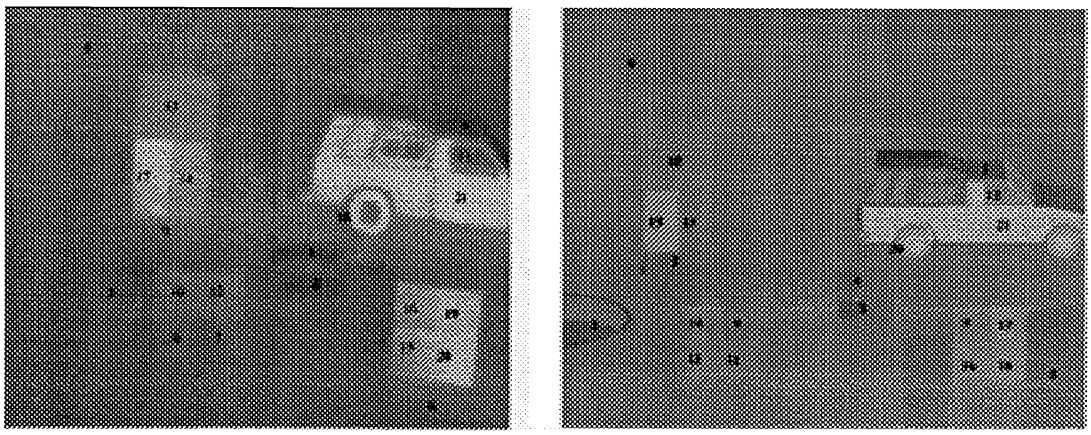

Fig. 4 LWIR truth (left) and DIRSIG images for 1200 hours are shown along with rank order levels.

correlation statistic runs from -1 to 1 and is interpreted in a fashion similar to the conventional correlation coefficients. Table 1 shows the data set for the 1200 hour images. The LWIR rank order values in the truth image correspond to the target locations indicated in Fig. 4. The values of the correlation statistic for the 24 hours of the June 1992 collection are shown in Table 2. These values are plotted in Fig. 5 as a function of TOD. The radiance range of the objects in each scene is also plotted for comparison.

The rank order correlation results indicate that for a wide range of targets, even with many of them quite similar, the DIRISIG model can quite faithfully mimic truth. The performance tends to be weakest when the overall scene contrast (apparent temperature range) is low which often coincides with thermal crossover times (i.e., when

Table 1
LWIR Rank Order Data from 1200 Hours GMT (06/23/92) ρ = 0.8649

	Radiance W/m^2sr		Rank Order		
Target Name	**Truth**	**DIRSIG**	**Truth**	**DIRSIG**	**Squared Difference in Rank**
Asphalt	27.58	28.84	8	5	9
Car Door	40.43	34.90	21	21	0
Car Roof	25.89	26.12	4	1	9
Car Tire	34.54	33.44	18	20	4
Car Window	28.77	30.75	11	12	1
Concrete	25.42	27.25	2	3	1
Grass	27.00	29.47	5	6	1
House Roof	30.35	30.59	13	10	9
Left Side	33.39	32.83	17	19	4
P1	28.13	31.07	10	14	16
P2	28.88	30.27	12	8	16
P3	27.02	31.07	6	13	49
P4	27.41	30.75	7	11	16
P5	32.74	30.43	16	9	49
P6	35.59	31.87	19	17	4
P7	32.22	31.71	15	16	1
P8	36.31	32.19	20	18	4
Right Side	32.15	31.23	14	15	1
Sand	25.20	28.36	1	4	9
Spec Ref	28.09	29.63	9	7	4
Water	25.54	27.24	3	2	1

Table 2.
Rank Order Correlation Values

	Correlation Value	
Time (hrs)	8 - 14 μm	3 - 5 μm
0000	0.64	0.48
0100	0.58	0.65
0200	0.31	0.64
0300	0.39	0.77
0400	0.20	0.58
0500	0.30	0.54
0600	0.42	0.36
0700	0.38	0.30
0800	0.36	0.16
0900	0.40	0.26
1000	0.50	0.21
1100	0.40	0.39
1200	0.86	0.70
1300	0.91	0.80
1400	0.91	0.70
1500	0.87	0.70
1600	0.79	0.70
1700	0.84	0.79
1800	0.89	0.86
1900	0.88	0.75
2000	0.89	0.70
2100	0.70	0.52
2200	0.85	0.59
2300	0.75	0.47

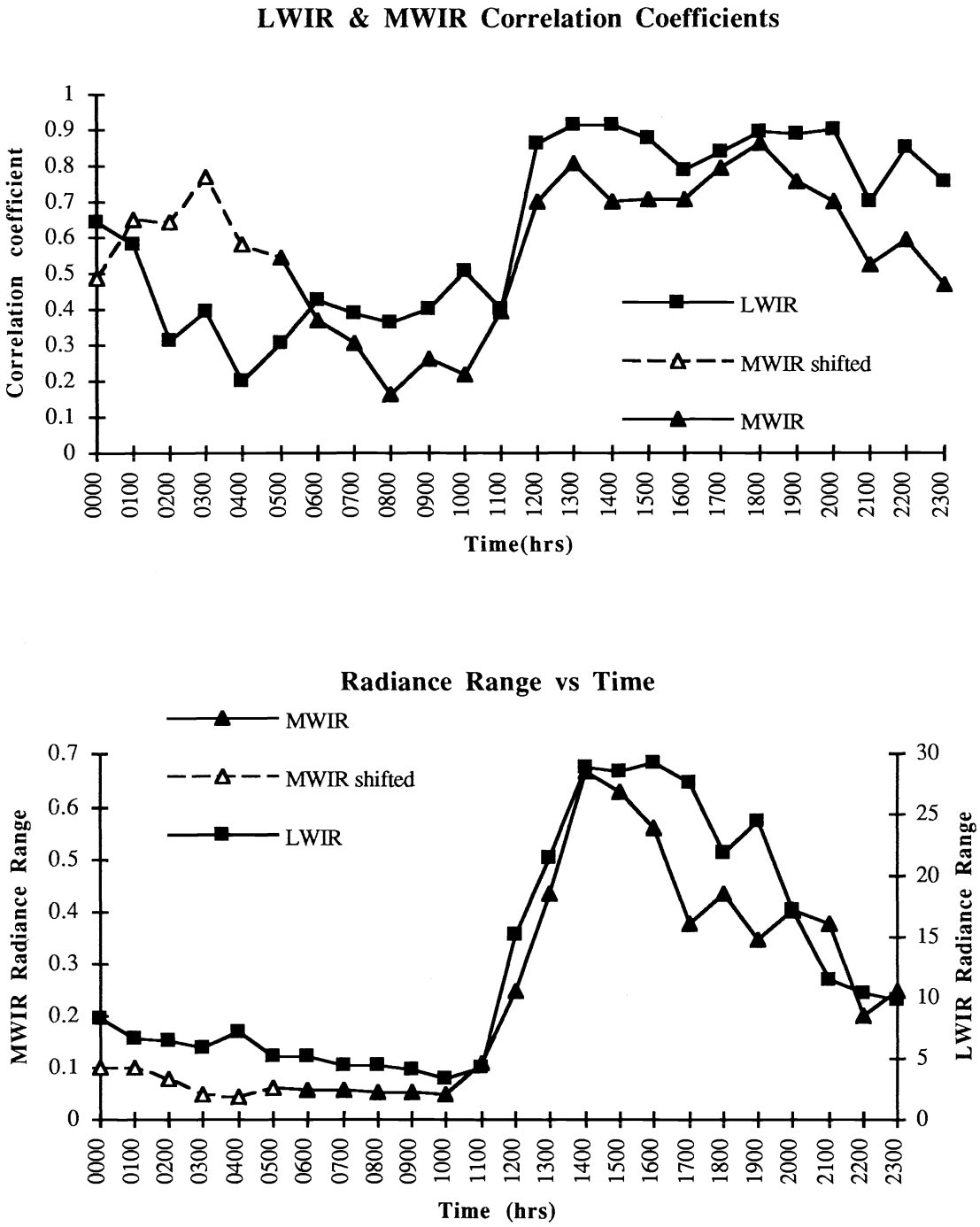

Fig 5. Rank order correlation and radiance range (w/m^2 sr) plotted as a function of Greenwich Mean Time (subtract 5 hours for local time) for the 24 hour validation experiment. The MWIR (shifted) is data taken from the next 24 hour cycle and wrapped around for clarity.

target-to-background contrast is reversing sign). However, even at these difficult times, DIRSIG preserves much of the rank order relationships. Ongoing work is focusing on ways to further improve the model to reduce the residual errors. One of the greatest sources of rank order error is associated with poor knowledge of object specific material properties. Simplified methods for deriving these properties are required for many operational uses of any SIG approach.

3.5 Comparison of Phenomenological Indicators

In examining actual thermal images, certain phenomena occur that are often not reproduced in synthetic images. In particular, interaction effects between targets and surround can be rather complex and/or subtle and slip through the cracks in a model. However, some of the effects are very important in controlling the brightness levels and, therefore, the appearance of a scene. In this section we will look briefly in a non-quantitative way at several of these phenomena to see if they are modeled by DIRSIG. This part of the study was not intended to validate how well these phenomena were modeled, but only to point out where significant phenomena might be overlooked. A good example of this is shown in Fig. 6 where a portion of a truth image, a DIRSIG validation image, and an upgraded DIRSIG image are shown. The truth image shows how the asphalt under the truck appears warmer. This is because the exposed area under the truck is smaller, reducing the radiational cooling to a cold sky. While the DIRSIG ray tracer computed the fraction of the exposed sky for use in the in-band radiometry calculations, this value was not used in the thermal model resulting in the validation image shown. Clearly the effects of this important phenomena are missing in the validation image. An upgraded version of DIRSIG incorporates this effect and correctly reproduces this phenomena as shown in the image in the center in Fig. 6.

Another limitation of DIRSIG at the time of validation is shown in Fig. 7. The shadows in the LWIR image are not due to reflection, but due to the heating of the surface based on how long it is exposed to sunlight. In the version of DIRSIG used in the validation, the sun-shadow history used in the thermal model was only updated every 15 minutes. This causes the stepped shadows shown in Fig. 7 due to the discrete motion of the sun (i.e., it jumps to a new position every 15 minutes). The deeper shadows are cooler, as they should be, but the abrupt temperature steps are unrealistic and can trigger an unexpected behavior in feature identification algorithms in addition to being visually distracting. The right hand image in Fig. 7 shows the result of an upgrade to DIRSIG to more closely approximate the continuous motion of the sun.

Other phenomena noted in the truth imagery and correctly modeled in DIRSIG were specular reflection of background objects, combined solar and thermal signatures in the MWIR, and differential heating of surfaces based on slope and azimuth angles. A major phenomena not modeled by DIRSIG was the local variation in brightness within an object. Thus the DIRSIG images in the validation are flat showing no texture variation within a material type. The upgraded version of DIRSIG includes material specific texture introduced as variations in each object's spatial and spectral reflectance. A detailed treatment of these upgrades is beyond the scope of this paper and will appear in future publications.

4. CONCLUSIONS AND RECOMMENDATIONS

This study pointed out the importance of using relative contrast as a metric for evaluation of the performance of SIG models. Depending on the application, either rank order or RMS error or both, may be the appropriate metric for evaluation of SIG imagery. The DIRSIG model was shown to perform quite effectively in terms of rank order correlation values for hourly comparisons over a 24 hour period. The degree of correlation was directly related to the scene's dynamic range with correlation values roughly halving ($\cong$0.8 to $\cong$0.4) with a factor of 6 decrease in radiance range. Much of the residual rank order error is due to errors in input material properties. Future efforts should focus heavily on developing improved data bases of material properties and development of methods for computing the material properties of scene elements.

The validation process included error propagation studies, RMS error, rank order correlation studies, and visual assessment for phenomenological features. It demonstrated a baseline performance level for Version 2.2 of DIRSIG and demonstrated a number of procedures that can be used in validation and comparison of SIG models. One of the primary goals of this effort was to identify limitations of the overall modeling process so that future R&D efforts could be more effectively focused. This goal was met by using the quantitative error analysis to point out the

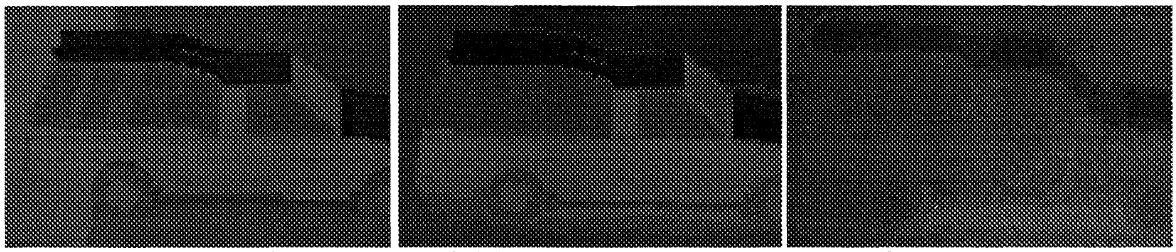

Fig. 6 Original validation image (left), DIRSIG image after improvements to account for radiation exchange in the thermal model (center), and LWIR "truth" image (right).

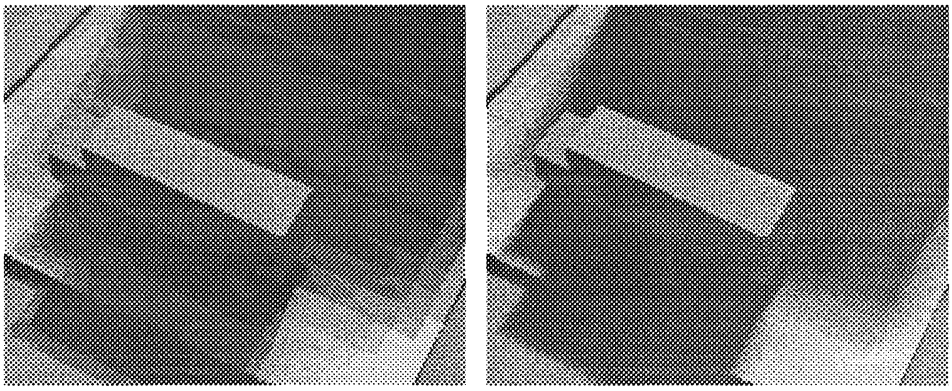

Fig. 7 Original DIRSIG image showing stepped shadows (left) and updated DIRSIG image when more nearly continuous solar motion is modeled (right).

major sources of error in the current model and to quantify the magnitude of the errors throughout the modeling process. The phenomenological assessment pointed out parameters and relationships that were not yet treated in the SIG model but which had important visual manifestations. Ongoing improvements to the DIRSIG model have been strongly focused as a result of the validation and error analysis study. We strongly recommend that other model developers consider similar quantitative validation studies an integral part of the validation process. Furthermore, we believe that development of objective validation metrics and testing methods should be encouraged and supported by the SIG community so that tradeoffs between different SIG modeling approaches can be effectively performed. We should point out that the emphasis here has been on tests of fidelity, particularly radiometric fidelity. We believe that additional metrics associated with operational issues such as ease of scene generation, run time, etc. should be developed and used in evaluation of SIG models.

5. ACKNOWLEDGMENTS

The authors would like to thank Ms. Carolyn Kitchen for preparation of this paper and Mr. Frank Tantalo for his work on data reduction. We also gratefully acknowledge the support of the Central Intelligence Agency for their support of this effort.

6. REFERENCES

1. Schott, J. R., R. Raqueno, C. Salvaggio, "Incorporation of a Time-Dependent Thermodynamic Model and a Radiation Propagation Model into Infrared Three-Dimensional Synthetic Image Generation", *Optical Engineering*, Vol. 37, No. 7, 1505, July 1992.

2. Mason, J.E., Schott, J.R., and Rankin-Parobek, D., "Validation analysis of the thermal and radiometric integrity of RIT's synthetic image generation model, DIRSIG," Presented at the SPIE Thermosense XVI, Orlando, April 1994

3. Schott, J. R., J. E. Mason, C. Salvaggio, J. D. Sirianni, R. A. Rose, E. O. Kulp, D. K. Rankin, "DIRSIG - Digital imaging and remote sensing image generation model: description, enhancements, and validation," RIT/DIRS Report 92/93-51-146, July, 1993.

4. DCS Corporation, "AIRSIM thermal signature and prediction analysis tool definition and analysis of object inputs," DCS Technical Note 9090-002-004, DCS Corporation, December, 1990.

5. DCS Corporation, "AIRSIM thermal signature prediction and analysis tool model assumptions and analytical foundations," DCS Technical Note 9090-002-001 DCS Corporation, December, 1991.

6. Berk, A., L.S. Bernstein and D. C. Robertson, "MODTRAN: A moderate resolution model for LOWTRAN 7," GL-TR-89-0122, Spectral Sciences Inc., April, 1989.

7. Kneizys, F. X., E. P. Shettle, L. W. Abreu, J. H. Chetwynd, G. P. Anderson, W. O. Gallery, J. E. A. Selby and S. A. Clough, "Users Guide to LOWTRAN7," AFGL-TR-88-0177, Environmental Research Papers, No. 1010, Air Force Geophysics Laboratory, Optical/Infrared Technology Division, Hanscom AFB, MD, December, 1988.

8. Fujino, S. T., Miyoshi, M. Yokoh and T. Kitahara, "Mitsubishi thermal imager using the 512x512 PtSi focal plane array," *Proceedings of the SPIE, Infrared Technology*, Vol. 1157, San Diego, CA, 1989.

9. Kendall, M.G., Rank Correlation Methods, Hafner Publishing Company, New York, 1962.