# A Standard Data Set for Performance Analysis of Advanced IR image processing techniques

A. Robert Weiß*[a], Uwe Adomeit [a], Philippe Chevalier[b], Stéphane Landeau[c], Piet Bijl[d], Frédéric Champagnat[e], Judith Dijk[f], Benjamin Göhler[a], Stefano Landini[g], Joseph P. Reynolds[h], Leslie N. Smith[i]

[a]Fraunhofer IOSB, Gutleuthausstr. 1, 76275 Ettlingen, Germany; [b]DGA Techniques Terrestres, Rocade Est – Echangeur de Guerry, 18021 Bourges Cedex, France; [c]DGA Ingenierie des Projets, 7-9 rue des Mathurins, 92221 Bagneux Cedex, France; [d]TNO BSS Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands; [e]ONERA DTIM, Chemin de la Hunière, 91761 Palaiseau, France; [f]TNO, Oude Waalsdorperweg 63, 2509 JG The Hague, The Netherlands; [g]Selex Galileo, Via Einstein, 35, 50013 Campi Bisenzio, Italy; [h]US Army Night Vision & Electronic Sensors Directorate, 10221 Burbeck Road, Fort Belvoir, VA 22060-5806, USA; [i]US Naval Research Laboratory, Applied Optics Branch, 4555 Overlook Avenue, SW, Washington, DC 20375-5320, USA

## ABSTRACT

Modern IR cameras are increasingly equipped with built-in advanced (often non-linear) image and signal processing algorithms (like fusion, super-resolution, dynamic range compression etc.) which can tremendously influence performance characteristics. Traditional approaches to range performance modeling are of limited use for these types of equipment. Several groups have tried to overcome this problem by producing a variety of imagery to assess the impact of advanced signal and image processing. Mostly, this data was taken from classified targets and/ or using classified imager and is thus not suitable for comparison studies between different groups from government, industry and universities. To ameliorate this situation, NATO SET-140 has undertaken a systematic measurement campaign at the DGA technical proving ground in Angers, France, to produce an openly distributable data set suitable for the assessment of fusion, super-resolution, local contrast enhancement, dynamic range compression and image-based NUC algorithm performance. The imagery was recorded for different target / background settings, camera and/or object movements and temperature contrasts. MWIR, LWIR and Dual-band cameras were used for recording and were also thoroughly characterized in the lab. We present a selection of the data set together with examples of their use in the assessment of super-resolution and contrast enhancement algorithms.

**Keywords:** thermal imaging, super-resolution, contrast enhancement, non-uniformity correction

## 1. INTRODUCTION

The increasing dominance of staring, fully digitized and highly sensitive thermal imaging systems during the last decades was accompanied by the development of increasingly sophisticated signal and image processing algorithms. While online processing in analog or semi-analog system was mostly limited to gain-and-level adjustment, boosting and analog filtering, digital systems enabled comparatively uncomplicated employment of global and local operations on both the spatial and thermal resolution of a thermal imaging system. Sometimes it cannot be avoided to reduce the actual information supplied by the system to a human-presentable form, notably in dynamic range compression, where thermal imaging data supplied with 12, 14 or even 16 bit resolution has to be scaled down in a suitable way to be presented on an 8 bit gray-scale display.

---

*robert.weiss@iosb.fraunhofer.de; phone 49 7243 992 355; fax 49 7243 992 299; www.iosb.fraunhofer.de

In this paper, the term "advanced image processing" refers to techniques that alter either the spatial or the intensity information content of thermal imaging data in order to increase usability, utility and accessibility of the data for human observers in (generally) a non-linear way. Examples for such algorithms are image super-resolution for undersampled systems, (local or global) contrast enhancement, dynamic range compression, turbulence mitigation, scene-based non-uniformity correction, and inter-band fusion.

It should be noted that many of these techniques require considerable computational resources and are thus often still only used for offline post-processing of images. It is a safe assumption, however, that the increasing power of embedded processors will soon enable built-in sophisticated image processing for thermal imaging systems, leading to a necessity to find an objective way for performance assessment and system comparison.
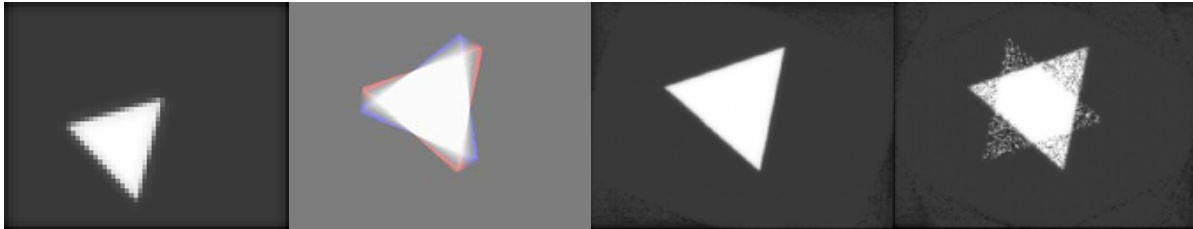


Figure 1. Super-resolution simulation. Single input image (left), rotational motion used as input to algorithm (center left), successful reconstruction (center right), reconstruction with periodicity failure for Fourier based method (right)

Traditional approaches to analytical modeling, like those used in the software packages NVThermIP [1] and TRM [2], are of limited use for several reasons. Locally operating algorithms violate the implicit assumption of isoplanacy for the pre- and post-sampling modeling chain; additionally, the actual performance of an algorithm can be strongly dependent on actual scene content. Likewise, the performance of super-resolution algorithms depends strongly on the content of the observed scene.

Laboratory assessment of many of these technologies likewise proves difficult; periodicities of standard test targets might confuse image registration and thus super-resolution techniques depending on shape and movement as depicted in Fig. 1. For contrast enhancement and image fusion, again the strong scene content performance dependence along with locally effective algorithms may render lab testing inefficient.

Other than live testing cameras and algorithms in field experiments, the only possibility for assessment then resides with the use of recorded pristine field data that is artificially altered to account for varying temporal and spatial noise characteristics, the sudden appearance of hot spots, etc. For specific algorithms and systems such measurements in conjunction with observer experiments have been done (see e.g., [3]). It is also possible to use image quality metrics like SSIM [4] to assess results for algorithm performance comparison purposes. While this approach is effective, comparison of algorithms developed by different institutions, companies, and universities is difficult due to the lack of a standard data set. Sometimes that data sets used, are partially classified and thus not publicly available. Other data sets are created on an as-needed basis, often geared towards a specific piece of equipment or research topic.

The NATO SET-140 / RTG–076 group on "Super-Resolution and Advanced Signal Processing for EO/IR Sensor Sensor Performance" has been working on solutions to the problems mentioned above since 2008. Little inroad could be made on analytical modeling and laboratory assessment of the above-mentioned image processing techniques due to the scene-dependence problem. We decided, however, to address the lack of a freely available standard data set for thermal imager image processing assessment for super-resolution (SR), contrast enhancement (CE), dynamic range compression (DRC), scene based non-uniformity correction (SB-NUC) and MWIR/LWIR fusion. For that purpose, a one-week measurement campaign was carried out in April 2010 at the Angers proving ground of the French Direction Générale de l'Armement (DGA-ETAS).

Section 2 will describe the imaging equipment used for the measurement campaign and overview of the types of recorded data along with a short description of targets and scenarios. Then, Section 3 shows two examples of the use of the data for a comparative assessment of SR techniques and for an observer experiment that investigated the relative utility of different CE / DRC approaches to target recognition. Section 4 concludes the paper.

# 2. MEASUREMENT EQUIPMENT AND ENVIRONMENT

Measurement equipment and environments were chosen in order to cover a variety of different thermal imaging technologies, applications and situations, including platform and scenario movements and intentional degrading of imaging parameters.

## 2.1 Thermal imagers

Three single-band thermal imagers were used during for all measurement scenarios. Their characteristics are given in Table 1. As can be seen, cooled LWIR and MWIR sensors were used alongside an uncooled bolometer; the uncooled systems were intended to deliver pristine and precisely calibrated imagery that can be artificially degraded in order to assess the sensitivity of algorithms against different kinds of distortions.

Table 1. Overview of the parameters for the cameras used for the measurement campaign

|  | LWIR | MWIR | Bolometer | Dual-Band |
|---|---|---|---|---|
| Model | FLIR SC7300 | FLIR SC7600 | FLIR SC660 | IRCam Geminis 110k ML |
| Sensor Technology | CMT | InSb | Amorphous Si | QWIP |
| Spectral Bands [μm] | 7.7-9.3 | 3.6-4.9 | 7.5-13 | 4.3-5.3 (MWIR) 7.7-8.7 (LWIR) |
| Resolution [pxl] | 320x256 | 640x480 | 640x480 | 384x288x2 |
| Pitch [μm] | 30 | 15 | 25 | 40 |
| NETD | < 25 mK | < 25 mK | < 45 mK | < 50 mK (MWIR) < 40 mK (LWIR) |
| Bit depth | 14 | 14 | 14 | 14 |
| Optics | F/2 - 22° | F/3 - 22° | F/1 - 24° | F/2 - 9° |

The bolometer was included since the temporal behavior of this type of sensor is more difficult to reproduce from the imagery obtained by other thermal imagers; this is important for smearing and interlacing effects due to the relaxation time of the sensor elements and the rolling shutter often employed in bolometers. These effects can be especially difficult to model when the performance of image registration and/or super-resolution algorithms are not limited by temporal or spatial noise, but by the momentary point spread function resulting from platform motion and movement within the observed scenario (see e.g., [5]).

The bolometer and cooled camera were mounted on a pan-tilt head, which can be programmed with accurate movement commands (see Figure 2). Programmed movements used during the measurement campaign included circles of various sizes for super-resolution imagery as well as panning movement for both image mosaics and SB-NUC (for this purpose, calibration in the cameras was switched off). All cameras were regularly calibrated during the measurement process using FOV-flooding with a blackbody.

## 2.2 Dual-band camera

Imagery suitable for investigations of MWIR/LWIR fusion algorithms was obtained using a custom built IRCam camera using a sandwiched QWIP detector [6]; again, the main parameters of the imager can be found in Table 2. In contrast to the single-band imagers, the LOS of the dual-band camera was fixed during all recordings.

Many fusion algorithms require an absolute temperature calibration in both wavelength bands to work properly. To enable the use of such algorithms, a two-point correction of the camera was done regularly by blackbody FOV-flooding during a measurement series (see also Figure 2). The data obtained is thus given in absolute temperature values with a resolution of 14 bits.

Since the imager is built around a sandwiched QWIP detector, images in MWIR and LWIR are automatically registered and no additional steps are necessary for band fusion; the band images can also be used individually, however, if specific conditions and properties of QWIP-based imagery are to be investigated.
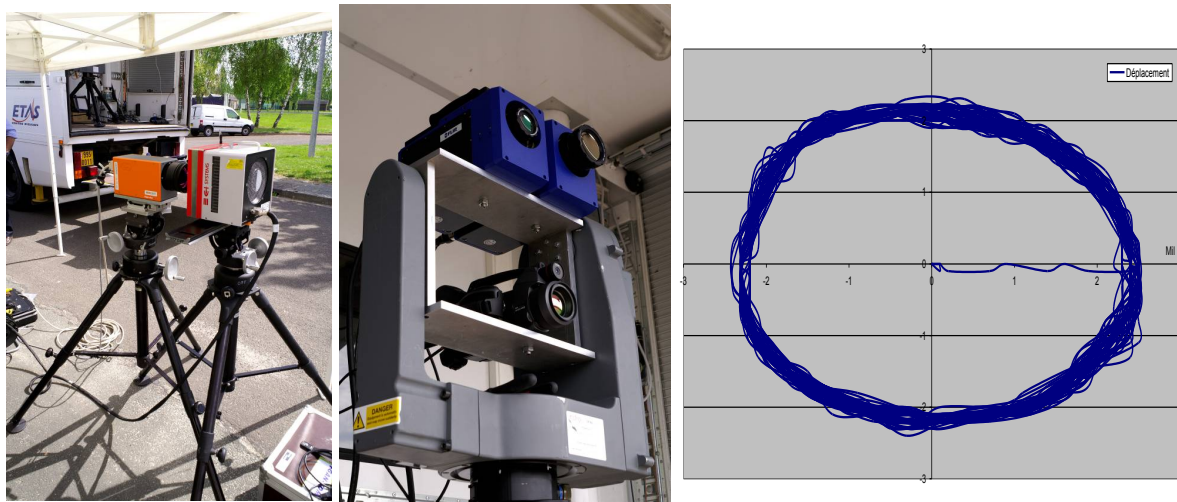


Figure 2. Dual-band camera with calibration blackbody in front of optics (left). Cooled MWIR / LWIR imagers and bolometer mounted on DGA ETAS' pan-tilt head (center). A measurement of one of the movement pattern commanded used during most super-resolution data collections (right).

## 2.3 Environments and Targets

In order to cover a variety of backdrops, distances and situations, four different environments available on the Angers proving ground were visited for the measurements; in the dataset they are called ROAD, URBAN, HILL and TOWER, respectively. Together with the targeted image processing techniques (SR, DRC/CE, Fusion, SB-NUC) and the cameras, many combinations of measurement sets would have been possible. Not all combinations, however, were used at all locations, as can be seen in Table 2.

Table 2. Overview of data sets recorded.

|  | LWIR | MWIR | Bolometer | Dual-Band | SR | DRC/CE | FUSION | SB-NUC |
|---|---|---|---|---|---|---|---|---|
| ROAD | x | x | x |  | x | x |  |  |
| URBAN | x | x | x | x |  | x | x |  |
| HILL | x | x | x | x |  | x | x | x |
| TOWER | x | x | x | x | x | x | x | x |

A short description of their general setup and the targets used is given below.

Figure 3. Main target of the ROAD dataset in the visible (left) and MWIR (right).

ROAD

The measurements of the ROAD scenario were carried out on a large tarmac area near the center of the proving ground. The distance from the sensor setup to the targets was usually around 50 m. Figure 3 shows a visual and thermal images of the main target used, a truck with a variety of standard targets and calibration equipment:

- Two 1 m x 1 m blackbodies on the back of the truck with 4 bar pattern targets of decreasing size. Generally, at least the two bottom rows of the rightmost target were not resolved and are thus suitable for the assessment of SR algorithms.

- A 1.2 m x 1.2 m tilted edge target for MTF measurements.

- A 0.5 m x 0.5 m Triangle-orientation-discrimination (TOD) target to the immediate right of the truck; again, the bottom rows of the target were calculated to generally be unresolved without SR.

- Two stacked blackbodies set to different temperatures for calibration purposes.

Data sets with commanded circular movements intended for SR and stationary sets for CE/DRC were recorded with all single-band cameras. The ROAD imagery of the truck can be regarded as the mainstay and baseline of our dataset since quantitative research can be done comparatively quickly using the available standard targets in the FOV.

In addition to the truck, a variety of other scenarios were recorded for the ROAD dataset:

- Static and dynamic images of people holding handguns (CE/DRC)

- Dynamic scenarios of increasing complexity with a car driving in circles and people riding a bicycle around a stationary car (SR)
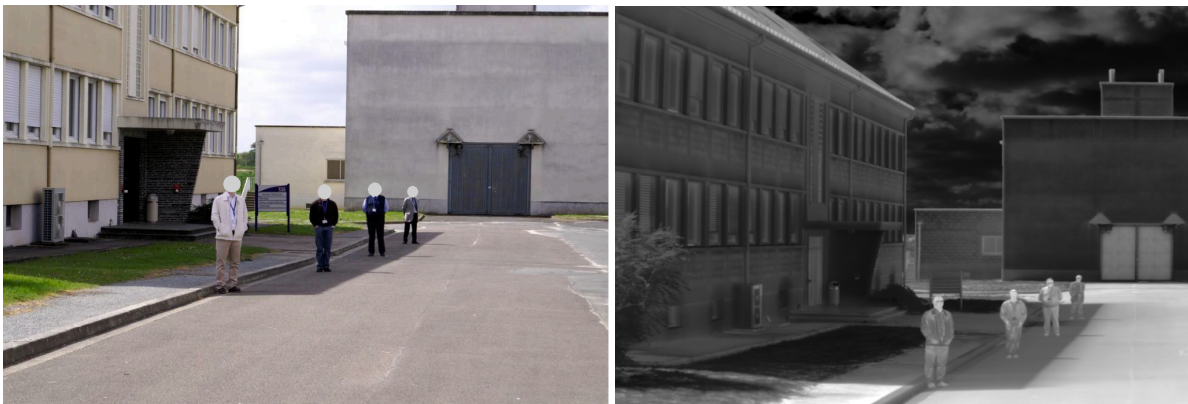


Figure 4. View from the URBAN dataset measurement position in the visual (left) and MWIR (right).

URBAN

The URBAN dataset was recorded close to the entrance of the proving ground on a short strip of road with an office building to the left and a storage area as a backdrop (see Figure 4). Data was recorded with all available imagers for CE/DRC and fusion assessment purposes only.

The URBAN scenario was intended as a stand-in for environments with comparatively low-clutter and short-distances, and should facilitate e.g., the pairwise comparison of DRC and fusion algorithms when used for urban surveillance thermal imagers. The distance from the targets to the measurement equipment varied from a few meters to about 30 m.

The following targets and scenarios were recorded:

- People standing at increasing distance in various lighting conditions (static).

- People moving in random directions (dynamic).

- Car approaching the sensor location with different velocities (dynamic).



Figure 4. View of the HILL measurement area in the visual (left) and MWIR (right). A hotspot blackbody can be seen a in the background.

HILL

The HILL dataset was recorded with the goal to assess fusion and CE/DRC algorithms in environments with intermediate distances and high clutter (see Figure 4). It also included a hot-spot blackbody in all images that can be masked and thus be used to investigate the reaction of CE algorithms to the sudden appearance of small, hot, targets in the FOV.

All cameras were used for the recordings with measurement distances varying between 100 and 300 m. The following scenarios were recorded:

- Five persons in varying configurations at the tree line and on the sand; between 3 and 5 persons are detectable in all recorded frames (see also Section 3).

- Stationary tank target on grass and on sand.

- Car driving in circles around a gravel track in front of the tree line (dynamic).

- Hot spot blackbody in the background of all images.

TOWER

The TOWER dataset was recorded from the Angers proving ground's measurement tower overlooking the test tracks (see Figure 5). At a height of approximately 33 m, the data is suitable to assess applications of image processing algorithm performance potentially in slow, low-flying platforms like helicopters or UAVs.

A tank and several cars were positioned in the field of view, as was a hotspot blackbody in the back of the images. Given that in-flight calibration of thermal imaging equipment can be a significant problem in aerial applications, a great number of panning datasets were recorded where the built-in NUC was either switched off or its quality artificially degraded by setting the integration time to values different from the calibration settings.



Figure 5. View from the tower in the visual (left) and MWIR (right). A blackbody hotspot can be seen to the back of the image.

## 2.4 Other data

The data is supplied together with an extensive protocol detailing the contents of the database, the setup of the recorded scenarios and the calibration files relevant for a given measurements. Additionally, the following data is available:

- Meteorological data recorded close to the measurement locations for all scenarios

- Commanded and measured movement patterns for the single-band camera pan-tilt head (where applicable)

- Visual band images of the measurement locations recorded in parallel with the thermal data set

# 3.  SAMPLE APPLICATIONS

The two main applications to be covered by our dataset are super-resolution and DRC/CE; thus, the following to sections detail two example uses of the set targeted at an assessment of these algorithms.

## 3.1 Dynamic Range Compression

A selection of data taken from the HILL measurement set was used to carry out an observer experiment at a US Army facility, comparing search performance using different dynamic range compression algorithms. Figure 6 shows an image mosaic of pristine images along with resulting images from the pristine set and a no-NUC dataset of the same scenario.

In order to create a variety of images to be shown to the observers, cropped images were created from the originals, containing with a between 0 and 5 visible persons in one image. Both the pristine image with artificial noise (TVH) and the No-NUC dataset were subjected to the following DRC/CE techniques:

- Min-Max-Stretch

- $2\sigma$-Stretch

- Contrast-limited adaptive histogram equalization (CLAHE)

resulting in a total of 708 images. These images were presented to 9 observers in random order. The observers were asked to mark any visible person in the image but were not told the maximum number. "No target" was a valid answer, and a 14 second decision time limit was enforced per selection.
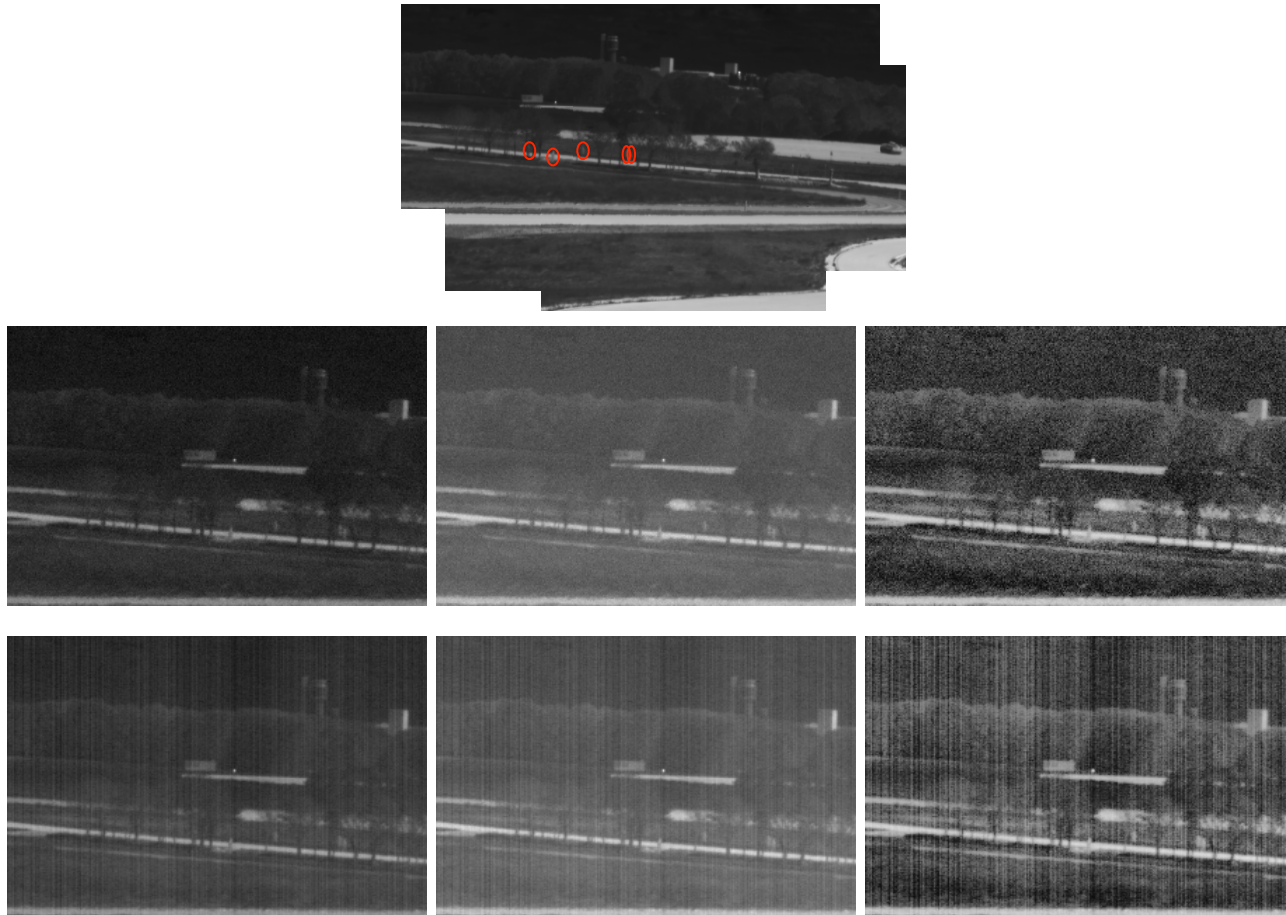
Figure 6. Input scenario image mosaic used to generate images for the observer experiments (top). Images taken from the cropped from the mosaic with artificially added noise (center row) scaled with Min-Max-stretch (left), 2σ-stretch (center) and CLAHE (right), respectively. Images taken with NUC switched off (bottom row) treated with the same algorithms.

The results of the experiment were then used to derive both the probability of detection *P(det)* and the false alarm rate FAR for the different CE/DRC algorithms. For this purpose, a box with a side length of 10 pixels surrounded each target in the FOV. Any click in a valid target area resulted in a valid detection, while any click outside was counted as a false alarm. The FAR for a single image was then calculated as the number of invalid detections divided by the number of 10x10 pixel boxes in the 300x200 pixel images.

As can be seen in Figure 7, there was virtually no performance difference between the different CE/DRC algorithms for a given sequence of images; curiously, however, the fixed pattern noise of the No-NUC images seemed to be less distracting for the observers than the artificial noise added to the pristine images, resulting in probabilities of distraction which were consistently 5 to 6 % greater. For the false alarm rate, however, no such effect can be seen.

## 3.2 Super-resolution

Another possible application of the data set is the assessment of super-resolution algorithms. The results described here were obtained by one of the authors of this paper at ONERA DTIM using a selected data from the ROAD measurement set.

The goal of this experiment was to investigate the convergence behavior of different SR approaches to a high-resolution image. For this purpose, a simple shift-and-add (S&A) algorithm was compared to the results using the same algorithm with additional Wiener filter boost (S&A+W) and an implementation of the Hardie SR algorithm [7] for different number of low-resolution input frames.
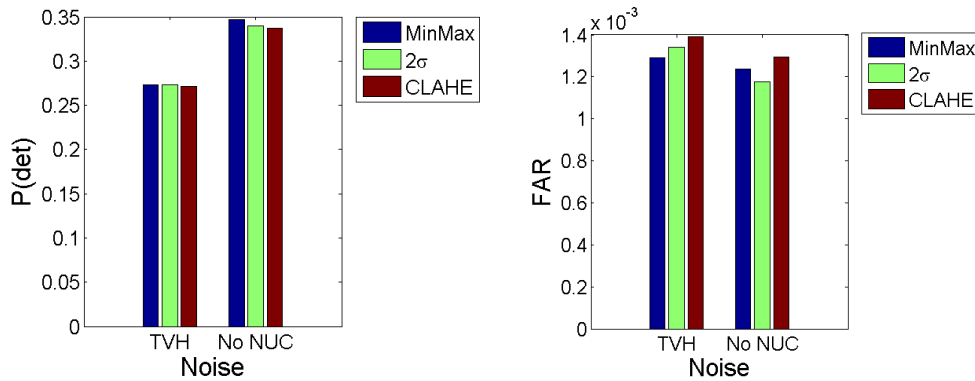
Figure 7. Observer experiment results for detection probability (left) and false alarm rate (right). TVH designates pristine images with artificially added noise.

Figure 8 shows both an input frame and the resulting frames using S&A, S&A+W and the Hardie reconstruction for 10, and 30 (different) input frames. As can be seen, both S&A+W and Hardie converge to a state where the third-to-last bar pattern can barely be resolved while the in fourth-to-last all four bars can be clearly distinguished. The lack of a boost for the simple S&A, which is implicit in Hardie's algorithm, probably explains that a pure S&A cannot compete with the two other approaches.

Generally, it turned out that the S&A+W did reach similar results as Hardie but needed more frames to be integrated until convergence.
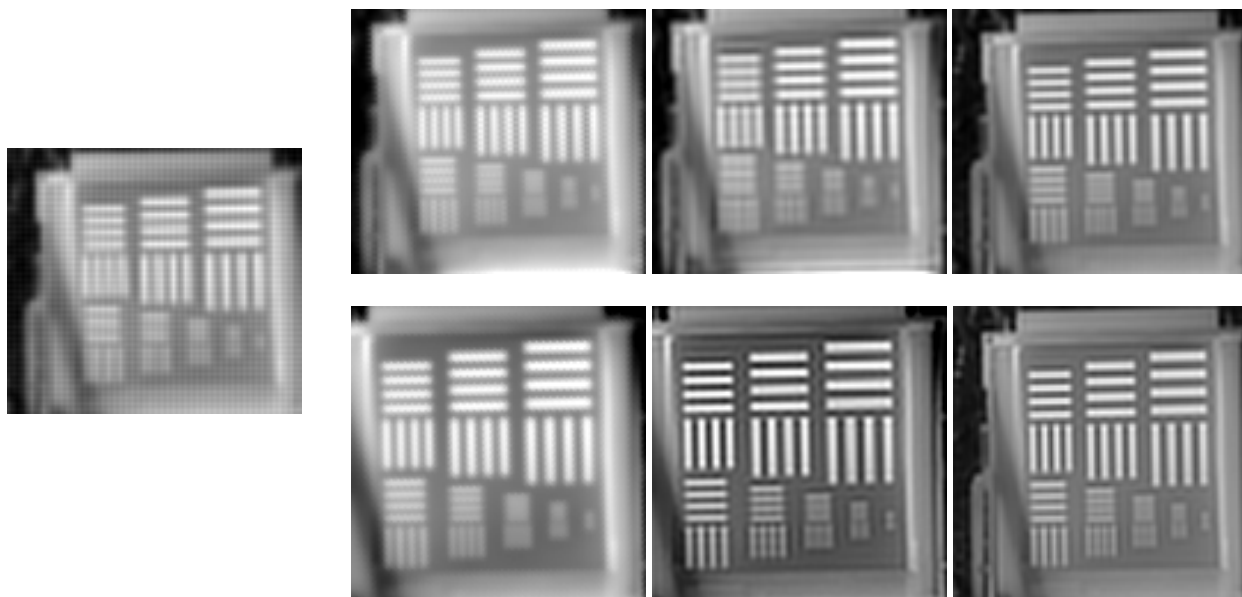


Figure 8. Single input frame (left). Super-resolved images of the four-bar blackbody target for 10 (top) and 30 (bottom) input frames for Shift&Add (left), Shift&Add+Wiener (center) and Hardie (right), respectively.

## 4. CONCLUSION AND AVAILIABILTY OF DATA

We presented the rationale for the creation of a freely available dataset that can be used to assess advanced image processing algorithms for thermal imagers. The areas of application include but are not limited to super-resolution, dynamic range compression and contrast enhancement, image fusion and scene based non-uniformity correction. Two simple uses for the assessment of CE/DRC and SR were presented.

At the time of writing, our group is in the process of selecting a suitable subset of all recorded data that will be thoroughly documented and then be made available to the public; in the meantime, datasets can be obtained by contacting the authors of this paper.

## ACKNOWLEDGEMENTS

## REFERENCES

[1] Bradley L. Preece, Jeffrey T. Olson, Joseph P. Reynolds and Jonathan D. Fanning, "Improved noise model for the US Army sensor performance metric", Proc. SPIE 8014, 801406 (2011)

[2] Wolfgang Wittenstein and Raanan Gal, "TRM3 progress report", Proc. SPIE 4130, 292 (2000)

[3] Jonathan Fanning, Justin Miller, Jennifer Park, Gene Tener, Joseph Reynolds, Patrick O'Shea, Carl Halford and Ron Driggers, "IR system field performance with superresolution", Proc. SPIE 6543, 65430Z (2007)

[4] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Transactions on Image Processing 13, pp. 600-612 (2004)

[5] Bijl, Piet, "Visual Image Quality Assessment with sensor motion: Effect of Recording and Presentation Velocity", Applied Optics 49, pp. 343-349 (2010)

[6] U. Adomeit and R. Ebert, "IR-dual-band-camera demonstrator: experimental assessment, practical applications", Proc. SPIE 6737, 673703 (2007)

[7] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong and E. A. Watson, "High Resolution Image Reconstruction from a Sequence of Rotated and Translated Frames and its Applications to an Infrared Imaging System", Optical Engineering 37, pp. 247-260 (1998)