Systematic infrared image quality improvement using deep learning based techniques

Huaizhong Zhang*[a], Pablo Casaseca-de-la-Higuera[a], Chunbo Luo[b], Qi Wang[a], Matthew Kitchin[c], Andrew Parmley[c], Jesus Monge-Alvarez[a]
[a]Centre for Artificial Intelligence, Visual Communications and Networking, School of Computing, University of the West of Scotland, Paisley PA1 2BE, United Kingdom; [b]Dept. of Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QJ, United Kingdom; [c]Thales UK, Glasgow G51 4BZ, United Kingdom

ABSTRACT

Infrared thermography (IRT, or thermal video) uses thermographic cameras to detect and record radiation in the long-wavelength infrared range of the electromagnetic spectrum. It allows sensing environments beyond the visual perception limitations, and thus has been widely used in many civilian and military applications. Even though current thermal cameras are able to provide high resolution and bit-depth images, there are significant challenges to be addressed in specific applications such as poor contrast, low target signature resolution, etc. This paper addresses quality improvement in IRT images for object recognition. A systematic approach based on image bias correction and deep learning is proposed to increase target signature resolution and optimise the baseline quality of inputs for object recognition. Our main objective is to maximise the useful information on the object to be detected even when the number of pixels on target is adversely small. The experimental results show that our approach can significantly improve target resolution and thus helps making object recognition more efficient in automatic target detection/recognition systems (ATD/R).

Keywords: Super-Resolution Convolutional Neural Network (SRCNN), IRT, super-resolution, ADT/R

1. INTRODUCTION

Infrared thermography (IRT, or thermal video) has been widely used in civilian and military applications such as surveillance, night vision and tracking, weather forecast, firefighting, facility inspections, etc. for collecting high quality data that are beyond the human visual perception range. The exceptional capacity of IRT comes from its capability to detect and record radiation in the long-wavelength infrared range of the electromagnetic spectrum [1]. Recent advances in IRT cameras have significantly improved the resolution and bit-depth of thermal images, which are traditionally regarded inferior to visual images, thereby making IRT images suitable and widely used in the scenarios with high value targets. For this reason, automatic detection and recognition of these targets has raised increasing interest in both academia and industry [2].

Object detection and recognition in IRT images collected under real-world settings is still a challenging research topic in computer vision. Such images are usually acquired at a very long distance, leading to extremely low values of pixels-on-target. A further challenge comes from the nature of IRT imaging: if the temperature of the object of interest is similar to the background, the contrast will be severely compromised. These adverse effects emerge as significant obstacles that degrade the performance of automatic object detection/recognition (ATD/R) in IRT images and hinder the application in practice. Fig. 1 shows two real-world image examples where the targets (a people carrier in (a) and an estate car in (b)) bear low resolution and poor contrast, which lead to high probability of false alarms in our developed ATD/R system (See Section 2.4). The “people carrier” target in Fig.1(a) has very low resolution (14×8 pixels), which is barely visible from a distance away. The “estate car” target in Fig.1(b) is almost blended with the background.

This paper aims to address these image issues to improve their quality so that the performance of ATD/R performance will be increased. A systematic approach based on image restoration and deep learning is proposed to increase the target signature resolution and optimise the baseline quality of inputs for ATD/R. The key of this approach is to maximise useful information of the objects to be detected even when the number of pixels on-target is so small that a slight amount

Remote Sensing Technologies and Applications in Urban Environments, edited by Thilo Erbertseder, Thomas Esch, Nektarios Chrysoulakis, Proc. of SPIE Vol. 10008, 100080P
© 2016 SPIE · CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2242036

of noise or non-uniformities would significantly degrade the detection performance. Our approach is composed of the following two essential steps: 1) A pre-processing and correction stage, which compensates undesired effects from the camera transfer function and small- and large-scale non-uniformities; 2) target signature resolution enhancement based on super-resolution and deep learning.

Traditional super-resolution methods in literature adopt example-based strategies to learn priors so as to constrain the solution space [3]. Inspired by the exceptional feature learning and extraction of deep learning, most recent trends [3, 4] resort to deep learning based methods to jointly optimise all components of the objects rather than separately handle each of them. For example, a deep super-resolution Convolutional Neural Network (SRCNN) has been proposed and proved to achieve state-of-the-art image enhancement performance [3]. Comparing with traditional image super-resolution techniques, the SRCNN method adopts a fully connected CNN to map low resolution images to their higher resolution counterparts by learning the essential features of objects. This can significantly improve target signature resolution, which is essential to ATD/R.

In order to apply SRCNN in our ATD/R system, this project collects a large set of real-world IRT data using a high-quality thermal camera (Thales Catherine MP Camera [5]), and generates the first thermal dataset suitable for training deep learning models. To the best of our knowledge, this is the first work applying SRCNN for thermal image enhancement in order to improve ATD/R performance. We implement the whole image restoration system and test its impact to ATD/R using two widely used metrics: Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [6]. Experimental results show that our proposed methodology can significantly improve image resolution and thus leads to more effective object detection and recognition for thermal images.

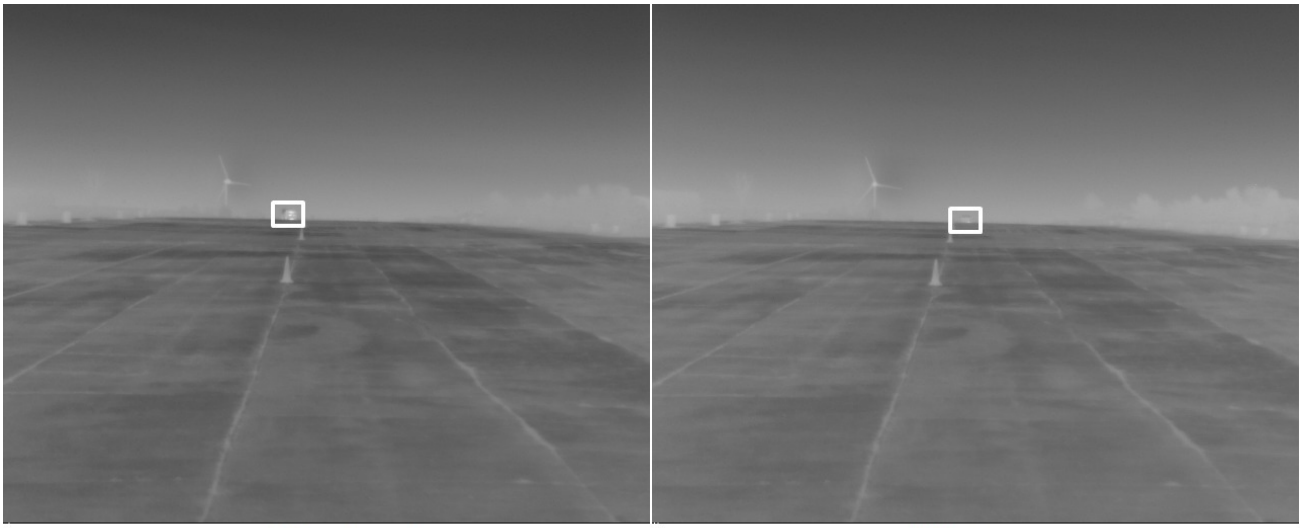

Figure 1. Two IRT image examples. The vehicle objects are highlighted in the bounding box.

2. DATA AND METHODOLOGY

2.1 Data acquisition and preparation

We firstly acquired raw images from surveillance video clips taken by Thales's Catherine MP LWIR camera, which is a specialised thermal camera using micro-scanning technology to combine the fields of resolution (640x512). Three types of vehicles, people carrier (nine-seat minibus), estate car, and van, were employed as targets. The targets were acquired using the camera wide field of view at 100m, 200m, 300m, 400m, 500m and 600m, and six groups of video clips were collected, each consisting of approximate 8,000 IRT images. Fig. 1 shows two example images with the objects "people carrier" and "estate car" from a video clip collected at 600m away, respectively.

2.2 Image preprocessing and camera bias correction

The video clips were acquired as raw data. Frames were extracted and converted to a suitable format to feed into the super-resolution algorithm.

IRT images have a small amount of random noise uncorrelated from pixel to pixel and small-scale non-uniformities from pixel to adjacent pixel. In our improved approach, the effects of these problems are significantly reduced and become trivial because the input image is pre-processed by subtracting the average image of the whole image data so as to greatly decrease image variation during model training and CNN calculation. This therefore helps make the performance of the super-resolution approach more efficient.

2.3 Image enhancement using SRCNN

In order to enhance the quality of the original IRT images, this paper adopts a recently proposed SRCNN algorithm for image enhancement [3], whose weights were trained using the raw IRT data randomly selected from the dataset. This section provides a brief introduction to SRCNN.

2.3.1 Introduction of the SRCNN method

The overall idea of super-resolution is that a low-resolution image $\boldsymbol{I}$ is upscaled to a new image $\boldsymbol{Y}$ using bicubic interpolation and then a mapping function F is employed to recover the high-resolution image $\boldsymbol{X}$ from $\boldsymbol{Y}$. To obtain F, a popular strategy is the following: first, generate patches from $\boldsymbol{Y}$ and represent them by a set of pre-trained bases, and thus obtain the feature maps of low-resolution images; second, a non-linear mapping is applied to the feature maps so that the representation of a high-resolution patch is generated; finally, the predicated high-resolution patches are averaged to produce the final full image. In SRCNN, these traditional operations are implemented by creating a three-layer CNN. The mapping F is conceptually obtained by a CNN framework, which consists of the following three operations (also shown in Fig. 2):

(Operation 1) Patch extraction and representation

This is the implementation of the first layer in Fig. 2. It can be described as an operation F_1:

$$F_1(\boldsymbol{Y})=\max(0, W_1 * Y + B_1), \tag{1}$$

where W_1 and B_1 are the filters and biases, respectively. W_1 applies n_1 convolutions on the input image, where the kernel size is $c \times f_1 \times f_1$, with c the image channel. The output includes n_1 feature maps. B_1 is an n_1-d vector associated with the filters.

(Operation 2) Non-linear mapping

The second layer in Fig. 2 is applied to implement the following operation:

$$F_2(\boldsymbol{Y})=\max(0, W_2 * F_1(\boldsymbol{Y}) + B_2) \tag{2}$$

where W_2 is a matrix of $n_1 \times 1 \times 1 \times n_2$ dimensions and B_2 is an n_2-d vector. Each of the outputs is an n_2-d vector that conceptually represents a high-resolution patch.

(Operation 3) Reconstruction

This convolutional layer in Fig. 2 produces the final high-resolution image by applying the following operation:

$$F(\boldsymbol{Y}) = W_3 * F_2(\boldsymbol{Y}) + B_3 \tag{3}$$

where W_3 is a matrix of $n_2 \times f_3 \times f_3 \times c$ dimensions, and B_3 is a c-d vector.

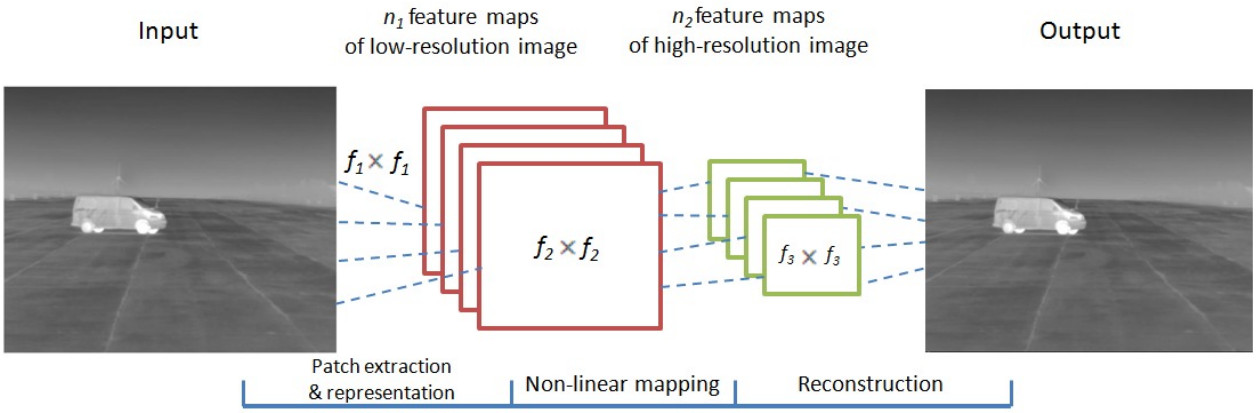

Figure 2. The block diagram of the super-resolution process using SRCNN. Here f_1, f_2, f_3 are the digital matrices. × denotes the convolutional operation.

2.3.2 Training the model weights with the acquired IRT images

The model weights, W_1, W_2, W_3, in Eqs. (1-3), are calculated by applying the standard stochastic gradient descent (SGD) algorithm. This is a back-propagation CNN process for the 3-layer CNN in Fig. 2.

The training set of 100 IRT images is randomly selected from our created IRT database. The following steps are performed to obtain the model weights.

- The ground truth images are prepared as 32x32-pixel sub-images randomly cropped from the training set.
- Low resolution images are pre-processed using Bicubic interpolation.
- The initial filter weights of each layer are generated by drawing randomly from a Gaussian distribution with zero mean and standard deviation 0.001. The learning rates are 0.0001 for the first two layers and 0.00001 for the last layer.

2.3.3 Applying the obtained model weights in SRCNN

In order to adapt the model to fit IRT data and enhance the images more effectively, we integrate the trained model weights into the SRCNN model instead of the original default weights. Thus, the collected IRT images can be improved properly according to the acquisition environment and modality properties in practice.

2.4 Validation

The validation of the proposed methodology is performed in our developed ATD/R system. We implement the system using the Caffe development platform [7], which supports the state-of-the-art Faster-RCNN technique for object detection and recognition [8]. Our specific validation focuses on vehicle surveillance based on IRT data.

We firstly use the improved images as the training set to train the CNN model in our developed ATD/R system. The training set consists of 2197 ground truth images, acquired via the camera's wide angle-of-view at the distances varying from 100m to 400m. This obtained new model is named as the super-resolution based model, which is used to conduct the object detection and recognition task in the ATD/R system. We then evaluate the model using a test dataset consisting of 575 ground truth images, acquired from the same camera view at different distances varying from 100m to 600m. The ATD/R system outputs the detected vehicles and types with the corresponding probabilities which can be interpreted as confidence values for detection/recognition. Finally, the object with the highest probability is treated as the recognised vehicle. The Receiver Operating Characteristic (ROC) curve is employed to assess the overall performance of the proposed approach.

3. RESULTS AND DISCUSSION

We employ Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) as performance metrics. Generally, a higher PSNR indicates that the image reconstruction contributes to higher image quality in terms of its difference with a reference image. SSIM can measure the similarity between the original image and the reconstructed version in a consistent way closer to human visual perception than traditional methods (e.g., PSNR) [6]. As shown in Table 1, our enhancement approach yields images that have higher average PSNR and SSIM in all experiments, where the PSNR and SSIM is obtained by averaging the gains over all test images of each class. Fig. 3 shows two examples of the SSIM index maps, which confirm that the reconstructed images can keep the structures of the images well, especially the vehicle structure is preserved in a good manner so that the objects can be detected with high accuracy in ATD/R.

Table 1. The results of PSNR (dB) and SSIM in the enhanced images. The values presented are the average over the test dataset.

	PSNR	SSIM
People Carrier	53.6	0.977
Estate Car	53.7	0.978
Van	52.1	0.975

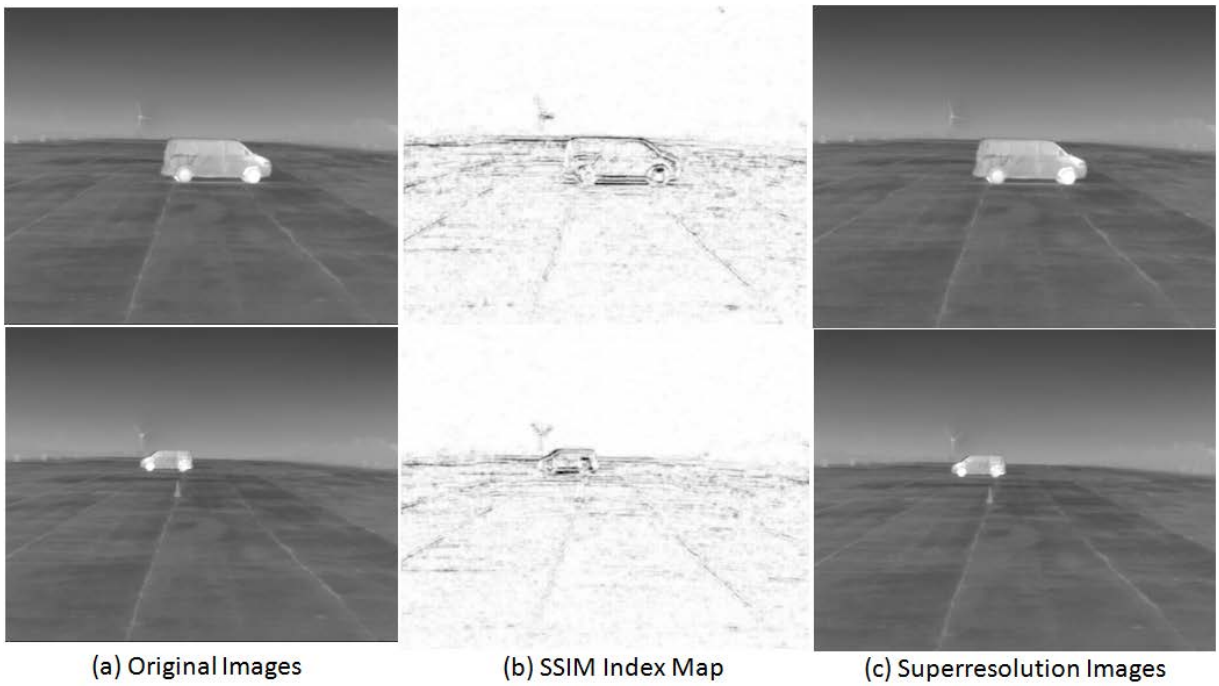

Figure 3. Examples of the generated super-resolution images (Col. (c)) and their SSIM Map (Col.(b)). The case of the 100m view (First row) is with PSNR=52.9dB and SSIM=0.9765; the case of the 200m view (Second row) is with PSNR=53.3 dB and SSIM=0.9769.

To evaluate the performance of improving image quality with the proposed approach, we use the enhanced training dataset to train the CNN model in the developed ATD/R system. The trained model is used to detect and recognise

objects in the test IRT dataset. The recognition results, represented by ROC graphs, are shown in Fig. 4. The ROC curve in blue shows the results from the original CNN model (raw model), which does not apply the proposed enhancement process. It can be seen that the Area Under Curve (AUC) obtained using the raw model is 0.92. The ROC curve in cyan is the obtained result from our proposed image enhancement method, which clearly illustrates better performance, with AUC reaching 0.96. It is worth noticing that, after adopting the proposed image enhancement approach, the true positive rates of the ATD/R system increase sharply during the false positive rates range of 0 to 0.2. We can interpret that the false recognitions are corrected from the obtained more accurate features through improving the quality of the training set.

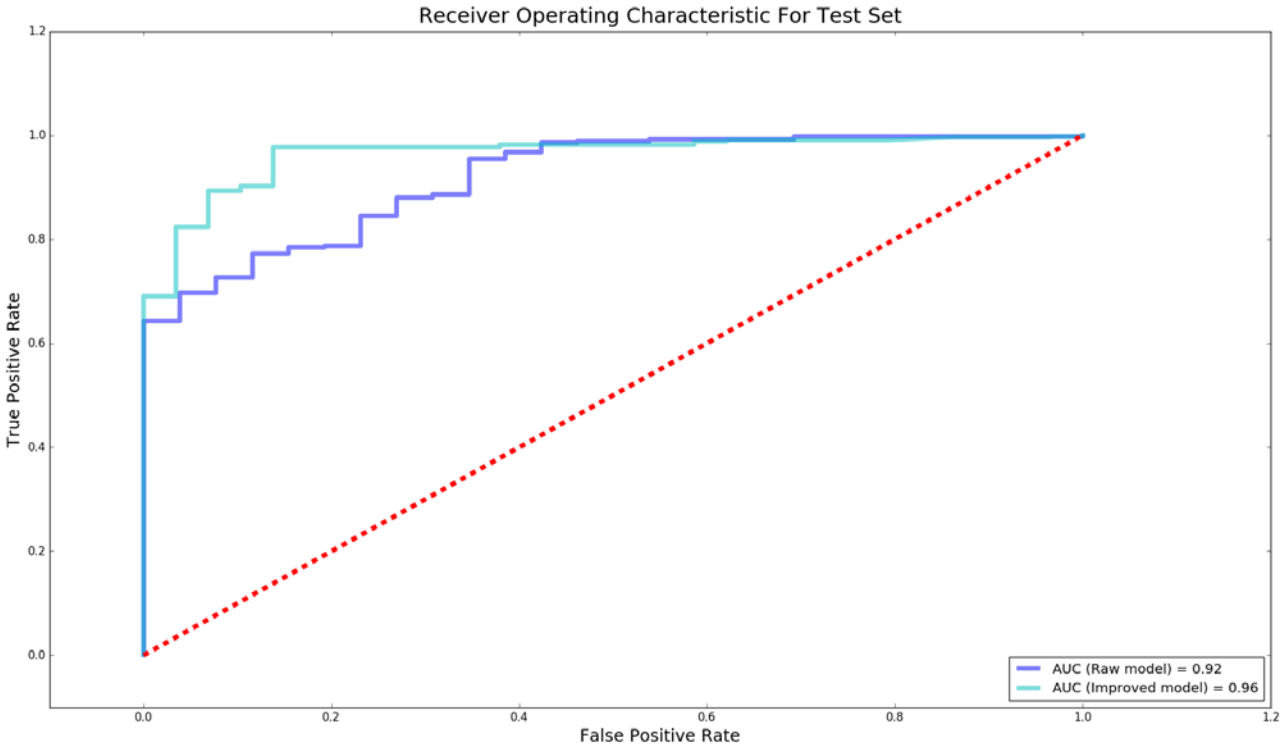

Figure 4. The ROC curves of the recognizing probabilities using our developed ATD/R system. Two models are used to generate the results: raw model and improved model. The ROC curve in blue denotes the raw model that is trained using the raw dataset; the ROC curve in cyan denotes the improved model using enhanced training images.

Table 2 specifically gives the detection confidences of the objects shown in Fig. 1, and provides some insight about the method. Before applying the proposed image enhancement method, the top three detection probabilities are: people carrier, 0.311699; estate car, 0.875782; and van, 0.0154172. Under these results, the target would be classified as an estate car since it has the highest detection confidence. However, if we check the ground truth, the target is actually a minibus. This error is clearly corrected by adopting the proposed methodology. The second row in Table 2 shows that, after image enhancement, the detection probability of "people carrier" is the highest among the three. Therefore, the ATD/R system can correctly recognise the target that was wrongly interpreted in the absence of image enhancement. Another example in Table 2, for the vehicle "estate car" in Fig. 1(b), exhibits detection confidence increasing from 0.205708 to 0.483660 after the data enhancement.

As shown by the ROCs, although the overall detection and recognition performance is improved, the small size of the objects at distant view can still cause some false positives. Due to the limited number of pixels on the small object signature being presented, the enhancement process may generate inaccurate feature information, leading to false positives. For example, a people carrier target is initially detected as a "people carrier" correctly with confidence of 0.729734 in the raw model; however, it is wrongly detected as an "estate car" with confidence of 0.772193 after the enhancement processing. This issue can be gradually improved by refining the training set in future work so that the ATD/R system can obtain more accurate feature information for recognition. In our system, such false positives

introduced by the enhancement process are rare and considerably outweighed by the significant performance gain. In fact, the ROCs show that the true positive ratios have been greatly improved after the enhancement process.

Table 2. The recognition results of Fig. 1 using raw training dataset and enhanced training dataset, respectively. The values presented denote the probabilities of each recognised vehicle type. The ground truths of Fig.1(a) and (b) are "People carrier" and "Estate Car", respectively.

	People Carrier	**Estate Car**	**Van**
Using raw training dataset (Fig. 1(a))	0.274491	**0.875782**	0.0154172
Using enhanced training dataset (Fig. 1(a))	**0.311699**	0.275520	0.0202263
Using raw training dataset (Fig. 1(b))	0.000419	**0.205708**	0.0014541
Using enhanced training dataset (Fig. 1(b))	0.001150	**0.48366**	0.0025878

4. CONCLUSION

This paper presents a systematic approach to improve IRT images for further automatic target detection and recognition. Our approach applies a CNN-based super-resolution method to improving the image quality, especially for the cases of long-distance view. Evaluation results show that the proposed approach has helped in improving the performance of the developed ATD/R system significantly, and thus validates the whole methodology. The study suggests that further work including developing advanced super-resolution methods, incorporating appropriate denoising techniques, and integrating the methods for creating a fully deployable system can be valuable extensions to the current method.

ACKNOWLEDGEMENTS

This work was funded by the Centre for Excellence for Sensor and Imaging System (CENSIS) and the Scottish Funding Council under the project "AALART. Thales-Challenge Low-pixel Automatic Target Detection and Recognition (ATD/ATR)", ref. CAF-0036. Thanks are also given to the Digital Health and Care Institute (DHI, project Smartcough-MacMasters), which partially supported Mr. Monge-Álvarez´s contribution, and to the Royal Society of Edinburgh and National Science Foundation of China for the funding associated to the project "Flood Detection and Monitoring using Hyperspectral Remote Sensing from Unmanned Aerial Vehicles", which partially covered Dr. Casaseca-de-la-Higuera´s, Dr. Luo´s, and Prof. Wang´s contribution.

REFERENCES

[1] Bagavathiappan, S., Lahiri, B., Saravanan, T., Philip, J. and Jayakumar T., "Infrared thermography for condition monitoring – A review", Infrared Physics & Technology, Vol. 60, pp. 35-55 (2013).
[2] Manolakis, D., Marden, D. and Shaw, G., "Hyperspectral image processing for automatic target detection applications," Lincoln Laboratory Journal, vol.14(1) (2003).
[3] Dong, C., Loy, C. C., He, K. and Tang, X., "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, pp. 295-307 (2016).
[4] Chen, S., Gong, H. and Li, C., "Super Resolution from a Single Image Based on Self Similarity," in Computational and Information Sciences (ICCIS), International Conference on, pp. 91-94 (2011).
[5] Craig, R. and Parsons, J., "Thermal Imaging for current D&S priorities," in Proceedings of the SPIE (2012).

[6] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. "Image quality assessment: from error visibility to structural similarity". IEEE Transactions on Image Processing. 13 (4): 600–612. doi:10.1109/TIP.2003.819861. ISSN 1057-7149 (2004).
[7] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T., “Caffe: Convolutional Architecture for Fast Feature Embedding”, arXiv preprint arXiv:1408.5093 (2014).
[8] Ren, S., He, K., Girshick, R., and Sun, J., “Faster {R-CNN}: Towards Real-Time Object Detection with Region Proposal Networks,” Advances in Neural Information Processing Systems (NIPS) (2015).