# Repeat multiview panchromatic super-resolution restoration using the UCL MAGiGAN system

Y. Tao, J.-P. Muller

Y. Tao, J.-P. Muller, "Repeat multiview panchromatic super-resolution restoration using the UCL MAGiGAN system," Proc. SPIE 10789, Image and Signal Processing for Remote Sensing XXIV, 1078903 (9 October 2018); doi: 10.1117/12.2500196

SPIE.

Event: SPIE Remote Sensing, 2018, Berlin, Germany

# Repeat multi-view panchromatic super-resolution restoration using the UCL MAGiGAN system

Y. Tao*[a], J-P. Muller[a]
[a]Imaging Group, Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey, RH5 6NT, United Kingdom

## ABSTRACT

High spatial resolution imaging data is always considered desirable in the field of remote sensing, particularly Earth observation. However, given the physical constraints of the imaging instruments themselves, one needs to be able to trade-off spatial resolution against launch mass as well as telecommunications bandwidth for transmitting data back to the Earth. In this paper, we present a newly developed super-resolution restoration system, called MAGiGAN, based on our original GPT-SRR system combined with deep learning image networks to be able to restore up to 4x higher resolution enhancement using multi-angle repeat images as input.

**Keywords:** Super-resolution restoration, earth observation, multi-angle, deep learning, generative adversarial network, MAGiGAN

## 1. INTRODUCTION

Very high spatial resolution imaging data is playing an increasing role in many commercial and scientific applications of Earth Observation (EO). However, this is tensioned against the realisation that as spatial resolution increases, so does spacecraft mass, driven by optical systems and power requirements until there comes a point at which the trade-off no longer works for any but the largest surveillance satellites. This suggests that even with future optical communications, satellite images are unlikely to be able to resolve features smaller than 25cm in the near future for any usable swath-width. The Earth's atmosphere also plays a significant role both in limiting the highest spatial resolution due to turbulence or light scattering from aerosols, particularly at large scattering angles, or clouds, which obscure most of the land surface most of the time. For commercial applications, whether in precision agriculture, forestry mapping, intra-urban intelligence, maritime tracking and detection, and monitoring of key sites at very high levels of details for defence and security, applications are always hampered by the size of the smallest object we can resolve from an orbital probe. In this paper, we describe a new paradigm to use multi-angle views of a surface both to clear atmospheric obscuration on the basis that these are unlikely to form in the same place at a different time and use repeat multi-angle views to resolve much smaller objects.

Previously within the EU FP-7 Planetary Robotics Vision Data Exploitation (PRoViDE) project (http://provide-space.eu), we developed a novel super-resolution algorithm called GPT-SRR [1] to restore distorted features from multi-angle observations using advanced feature and an area matcher based on least-squares correlation, Gotcha [2], and a segmented 4th order Partial Differential Equation (PDE) based Total Variation (TV) (Bouzari, SIViP, 2004) regularization approach. This technique was first demonstrated to resolve new surface information on individual rocks (diameter<150cm), rover tracks [3], and new evidence for the Beagle-2 lander using multi-angle repeat-pass 25cm NASA Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) images of the Martian surface [4].

More recently within the UK Space Agency CEOI SuperRes-EO project, we have further developed the SRR system using advanced machine learning algorithms. The new machine learning features are based on the Mutual shape adapted [5] Features from Accelerated Segment Test (O-FAST) [6] combined with Convolutional Neural Network (CNN) [7] feature matching, a Support Vector Machine (SVM) and Graph Cut (GC) based shadow modelling and removal [8], and the Generative Adversarial Network (GAN) [9] deep learning based super-resolution refinement. The new MSA-FAST-CNN-GPT-GAN (which we dub as "MAGiGAN") system not only retrieves subpixel information from multi-angle distorted features from the original GPT algorithm, but also uses the loss calculated from feature maps of the GAN network to replace the pixel wise difference based content loss of the original GPT algorithm to retrieve high texture

Image and Signal Processing for Remote Sensing XXIV, edited by Lorenzo Bruzzone, Francesca Bovolo, Proc. of SPIE Vol. 10789, 1078903 · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2500196

detail. The MAGiGAN system has been applied to stacks of 4m UrtheCast Corp. Deimos-2 (MS band) repeat-pass images over several experimental sites to produce SRR results with up to 4x resolution enhancement.

## 2. THE MAGIGAN SYSTEM

### 2.1 Summary of the original GPT-SRR workflow

The original implementation of the GPT-SRR system is shown schematically in Figure 1. In GPT-SRR, we take multi-angle Lower resolution (LR) images and a reference Orthorectified Image (ORI) as inputs. The workflow starts from MSA based Scale Invariant Feature Transform (SIFT) feature detection and the Gotcha process to calculate dense (at every pixel) sub-pixel correspondences for each input LR images with respect to the reference ORI. A high resolution (HR) image is initially estimated by placing the intensity values of the LR images on a HR grid. If a pixel in the HR grid does not have any correspondence from all LR inputs, this HR pixel will be interpolated from its neighbouring HR pixels. The HR grid is then optimised to the output SRR via a maximum a posteriori probability (MAP) process.

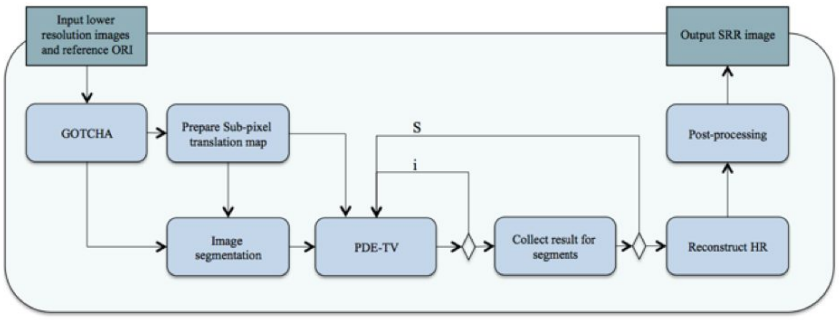



*Figure 1. Flow diagram of the original GPT-SRR system [1].*

The motion maps provide the initial degradation information in the similarity measurement term of the MAP estimation. LR images and the reference ORI are resized by the defined scaling factor and segmented into tiles according to a given threshold of the maximum difference of the magnitude of the distance of the motion vectors. Then the MAP process is resolved in the following steps: (1) for the same area, each tile (t) of each LR image (k) is projected with motion vector (matrix F), convolved with the first estimation of the Point Spread Function (PSF) (matrix H) which is assumed to be a small Gaussian kernel with various standard deviations according to the size of segments (S), down-sampled (matrix D) with the defined scaling factor (L); (2) upscale the degraded image and compared with its original LR image tile sequentially; (3) go back to (1) for the next image (k) until all images completed; (4) adding the transposed difference vector for the image tile (k, t); (5) adding the smoothness term and decomposing the TV regularization term with a 4th order PDE; (6) repeating from (1) using the further degraded LR image as inputs for the next steepest descent iteration until it converges; (7) collecting the HR result for this tile (t_i) and then going back to (1) for the next tile (t_i+1) until all segments (S) converge; (8) collecting the results for all HR segments (S) and reconstructig the full HR grid; (9) finally, a series of post-processing operations are performed based on the HR reconstruction including noise filtering and de-blurring.

A more detailed description of the method can be found in [1]. The GPT-SRR system is unique and optimised for planetary orbital image restoration, since (a) it not only uses marginal information from pixel shifting but also restores distorted features onto an ortho-rectified grid from comparatively large viewing angles (b) it uses a novel segmentation-based approach to restore different features separately; (c) applies our 5th generation adaptive least squares correlation region growing matcher (Gotcha), and a PDE solver based on a TV regularization approach to provide robust (noise resistant) restoration. The original GPT-SRR system has been applied to multiple 25cm resolution NASA MRO HiRISE images to restore 5cm-12.5cm near rover scale images (equivalent to rover Navcam imagery projected FoV at a range of ≥5m).

## 2.2 The MAGiGAN SRR system

There are three major differences between the EO imaging data and the Mars imaging data. Firstly, the Martian surface is usually static over many years, whereas the surface of Earth changes over different times of day and over different seasons. Secondly, the MRO HiRISE instrument captures images at a fixed local time which result in similarly shadowed regions, whereas EO imaging data always show different shadows and saturated bright pixels for the same area in different images. Thirdly, EO imaging data is generally noisier due to various atmospheric interferences including smoke, haze, and clouds. The new MAGiGAN SRR system aims at resolving these problems by producing denser initial feature correspondences on de-shadowed input images and applying a deep learning image network to further refine the SRR result.

The implementation of the MAGiGAN SRR system is shown schematically in Figure 2. There are several major changes in the feature matching and image segmentation algorithms compared to the original GPT-SRR system. Firstly. the MAGiGAN system uses a SVM and GC based shadow modelling and removal technique to minimise shading differences of the input LR images in order to obtain feature correspondences between shaded pixels and unshaded pixels for the same region. Note that this step is only for obtaining feature matching between shaded pixels and unshaded pixels in order to produce motion maps with less gaps in the Gotcha process at the next stage; the output SRR will keep the shading information from the reference ORI and is not devoid of shadows. The GC segmentation results from the shadow labelling stage are passed through to the tiling step to replace the naive segmentation attempts used in GPT-SRR. Secondly, the MAGiGAN system uses a novel MSA-FAST-CNN based feature matching to obtain much denser initial correspondences for region growing. In addition, a GAN network based single image SRR technique has been added to the last of the workflow steps to further enhance the output by training with a rich dataset of LR and HR images. A detailed description of the changes is given in the following subsections.

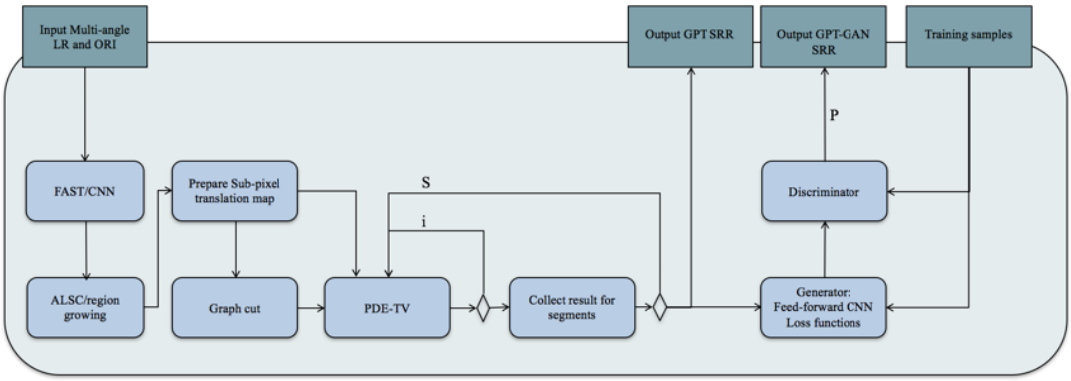



*Figure 2 Flow diagram of the MAGiGAN SRR system*

## 2.3 Image Segmentation and de-shadowing

Differences in shading effects may be treated as image content differences (like clouds) when matching multiple LR images and thus can produce motion maps with a lot of gaps. In GPT-SRR, gaps are interpolated using neighbouring pixels on the HR grid. Due to similar solar zenith angles and consequent orientations of shadows within the HiRISE images for the same area and less surface changes on the Martian surface, the gaps are generally minor for HiRISE SRR. However, EO images generally contain large amount of shadows, particularly in urban or forested areas with different orientations due to a range of solar azimuths away from the equator. Therefore, an area based de-shadowing process Figure 3 is applied in MAGiGAN. The first step is to segment the LR images based on their image content using the GC algorithm [10]. Segmented image patches for the same region from multiple LR images are then paired using normalised cross-correlation. If paired segments are found with the same illumination, they should be labelled with the same shadow notation (either shadow or non-shadow). If paired segments are found with different illumination, and one segment is much darker (at a given threshold) then the other one, the darker segment is labelled as a shadow. In order to increase the confidence of the shadow labelling, a pre-trained SVM [11] classifier is used to correct the shadow labels from the texture pairing approach. The SVM classifier is pre-trained using manually selected shadow and non-shadow segments from the GC results. Connected shadow patches are then grouped to one shadow segment and connected non-shadow

patches are grouped to one non-shadow segment. Finally, the illumination of the shadow segments is corrected using illumination statistics from the neighbouring non-shadowed pixels. Note that not all shadow segments are correctable, the intensity (texture) information may already have been lost during imaging. Such shadow segments will have a very low signal to noise ratio (SNR) after de-shadowing. This means that there will be no (or only a few) added feature correspondences for that region after the de-shadowing process. In which case, the de-shadowed segments should be reversed back to their original intensities.

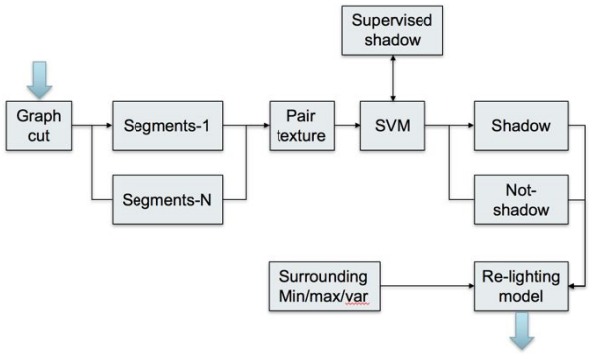



*Figure 3. Flow diagram of the de-shadow process*

The de-shadowed LR images are only used as metadata to provide seed feature points for the shadowed regions and will not be used in the follow-on processing stages. The segmented patches are passed through to the next processing stages for sub-segmentation based on the threshold of the maximum differences of the magnitude of the distance of the motion vectors. This is for further optimisation of the tiling process of the PDE-TV regularization step in order to restore different types of image content separately.

*Figure 4. Example of the original Deimos-2 image (PAN_L1C_20160923T105559) with shadow.*

*Figure 5. Example of the de-shadowed image.*

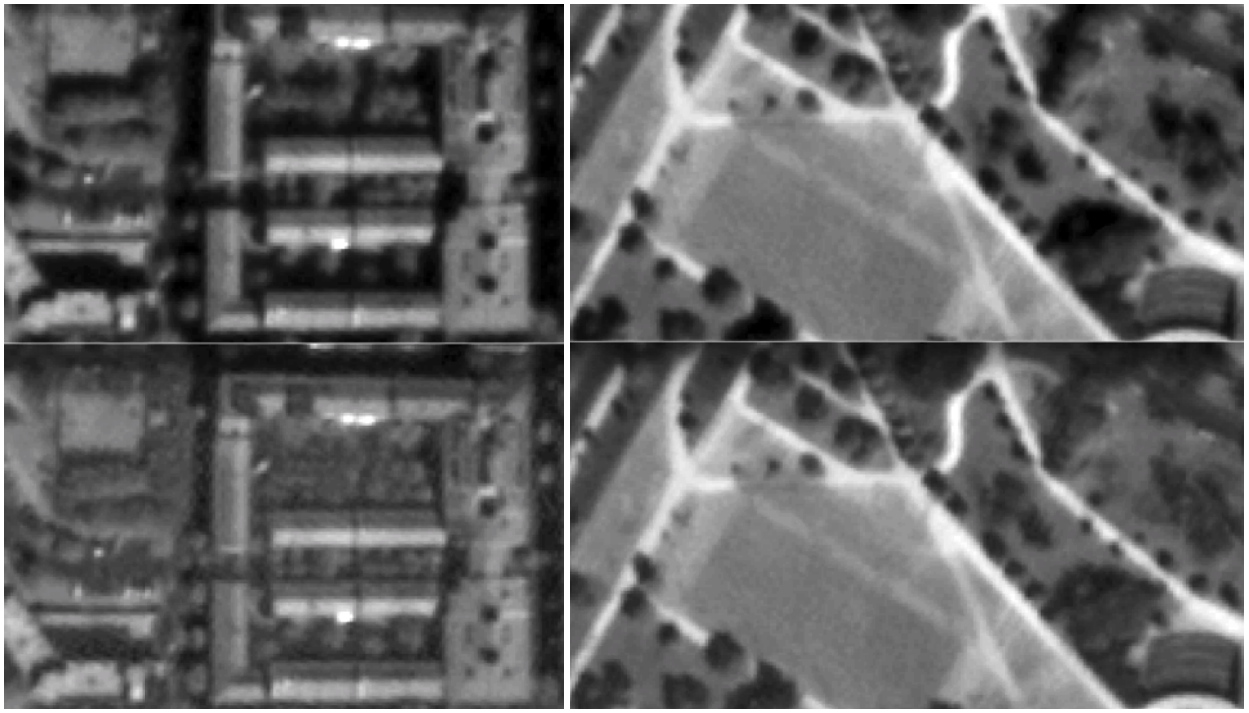


*Figure 6. A zoom-in view of the original Deimos-2 image (top) in comparison with the de-shadowed image (bottom) for building shadows (left) and tree shadows (right).*

Figure 4 shows an example of the original Deimos-2 image, in comparison with the de-shadowed Deimos-2 image shown in Figure 5. Figure 6 shows two zoomed-in comparisons of the de-shadowed results for building blocks and tree shadows.

### 2.4 Feature matching

Subpixel feature correspondences are used as seed points in the follow on ALSC and region growing stage to generate accurate motion maps of pixel correspondences for all LR inputs. An accurate and dense first estimation of the seed points is essential to the success of interpolating the initial HR grid. Ideally, the subpixel feature correspondences derived at this stage should be refined to an accuracy of 0.01 pixels and distributed evenly among different types of image content (e.g., building blocks, trees, road, flat regions, shaded regions, and saturated regions). The original GPT-SRR system uses a SIFT [12] based feature matching approach. In comparison, the MAGiGAN SRR system uses a FAST-CNN [6] [7] based feature matching approach to produce much denser initial feature correspondences (at a higher processing speed).

The FAST feature detector only considers a circle of 16 pixels around a local maxima point; if 12 (out of 16) pixels are all brighter or all darker by a given threshold from the centre point, then it is considered to be a feature point. These initial feature points are then refined using a pre-trained decision tree classifier (ID3 algorithm) to produce optimal choices of feature points. Adjacent feature points are compared according to their sum of absolute differences between the feature points and 16 surrounding pixel values. The adjacent feature points with a lower sum of absolute differences are discarded. The feature representation is trained with a CNN using a Maximally Stable Extremal Regions (MSER) detector which has been previously demonstrated in [7] to outperform all local descriptors including SIFT. In this work, we use an unsupervised CNN to obtain descriptors of patches around each FAST feature points. Firstly, at each scale, circular patches of 15 by 15 pixels around each FAST feature points are extracted. Secondly, a pre-trained CNN model consisting of 3 convolutional layers, proposed in [12], is used to extract descriptors from all patches. The extracted descriptors include output vectors from all 3 convolutional layers and a fully connected layer. The FAST-CNN method produces much denser feature correspondences compared to the SIFT (in GPT-SRR) and SURF methods [see Figures 7-9]. Finally, the MSA method described in [5] is used to refine each matched feature points to sub-pixel level. In MSA, the seed point locations and orientations from the FAST-CNN matched features are iteratively updated using forward and backward ALSC within a transformable elliptical window.

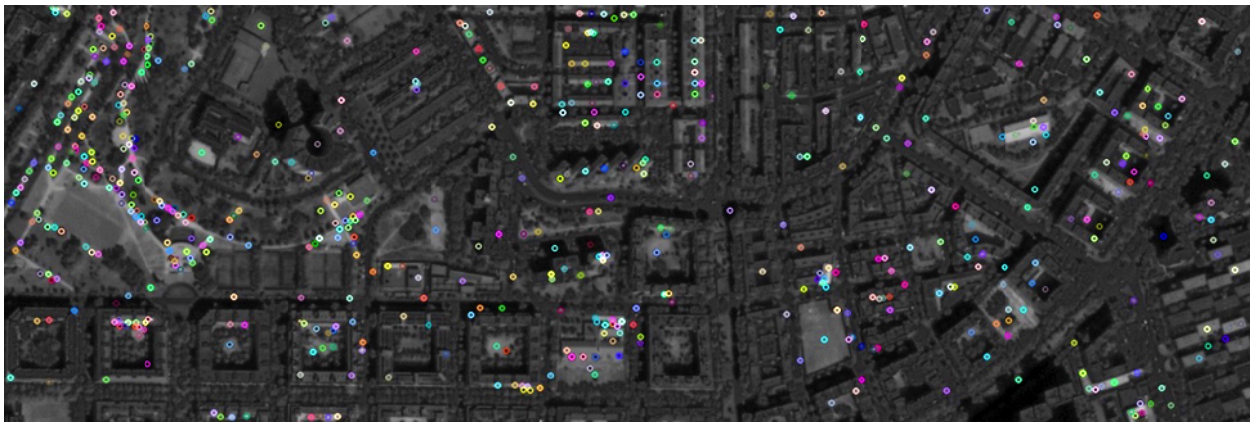


*Figure 7. Example of the matched seed points using MSA-SIFT (used in GPT-SRR) for a de-shadowed Deimos-2 image (PAN_L1C_20160923T105559) showing sparse distributions.*

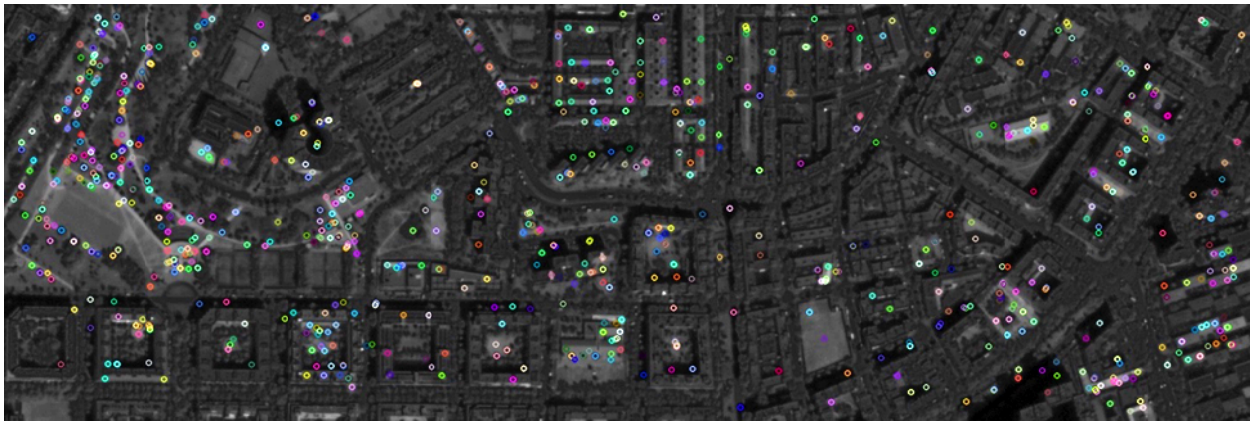


*Figure 8. Example of the matched seed points using MSA-SURF for the same de-shadowed Deimos-2 image (PAN_L1C_20160923T105559) showing sparse distributions*

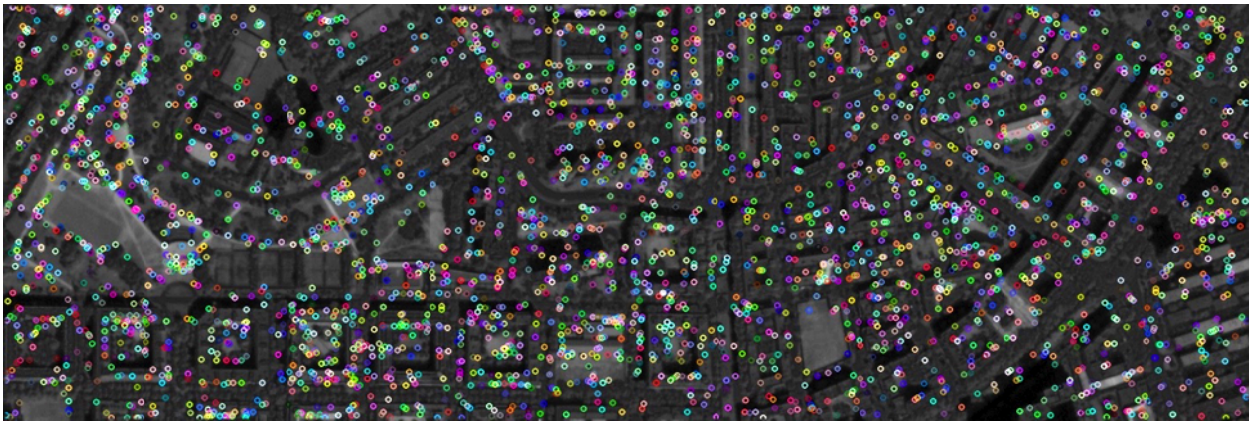


*Figure 9. Example of the matched seed points using MSA-FAST-CNN (used in MAGiGAN) for the same de-shadowed Deimos-2 image (PAN_L1C_20160923T105559) showing dense and more evenly distributed seed points.*

Figure 7 shows an example of the matched seed points using MSA-SIFT (in GPT-SRR) for a de-shadowed Deimos-2 image. Figure 8 shows an example of the matched seed points using MSA-SURF. Figure 9 shows the much denser and more evenly distributed matched seed points using MSA-FAST-CNN (in MAGiGAN). The dense initial seedpoint correspondences are essential to produce a more accurate motion map in the follow on Gotcha and PDE-TV process, which are further described in [1].

### 2.5 GAN based SRR refinement

GAN networks have been widely used in many generative tasks in computer vision, producing breakthrough improvements in performance compared to traditional image processing techniques. The original GPT-SRR system introduces information from multi-angle views, but the MSE based measurement (iteratively between HR and LR) does not generally retrieve high texture details as they are based on pixel wise differences. On the other hand, GAN uses the losses calculated from feature maps of the deep learning network to replace the MSE based content loss, and is therefore highly complementary to GPT in terms of restoring different features. In the MAGiGAN system, we have added a GAN single image SRR refinement step using the SRR output from MSA-FAST-CNN-GPT. GAN applies a deep network (Generator G) to generate high frequency textures that are highly similar to real images, in combination with an adversarial network (Discriminator D) to distinguish super-resolved images from real images. In this work, we use the GAN network described in [9]. The generator network uses B residual blocks consisting of 2 convolutional layers with small 3 by 3 kernels and 64 feature maps followed by batch-normalisation layers for each residual block. The discriminator network contains 8 convolutional layers with 3 by 3 filter kernels and 64 to 512 feature maps. The loss function for the generator consists of the content loss and adversarial loss. The loss function of the discriminator uses a typical GAN discriminator loss [13]. In MAGiGAN, we also applied the modifications suggested by [14] and [15] in order to balance the training strength between generator and discriminator. This includes removing the sigmoid activation from the discriminator network, not using the logarithm in the loss calculation of both generator and discriminator, constraining the weights to a constant range, and using stochastic gradient descent (SGD) to replace the momentum based optimiser.

*Figure 10. An example of a cropped sample of Deimos-2 4m MS green band image (DE2_11542_DE03; left) and 2m GAN single image SRR result (right).*

Figure 10 shows an example of the GAN single image SRR result (right) in comparison to the original Deimos-2 image (left).

## 3. RESULTS

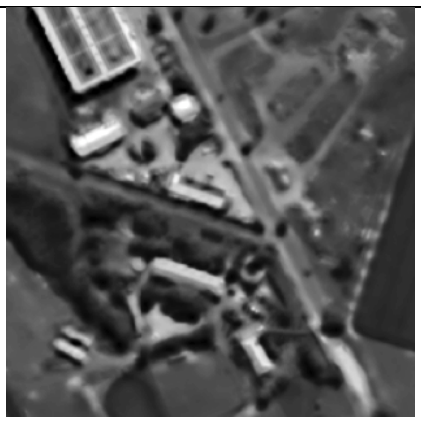
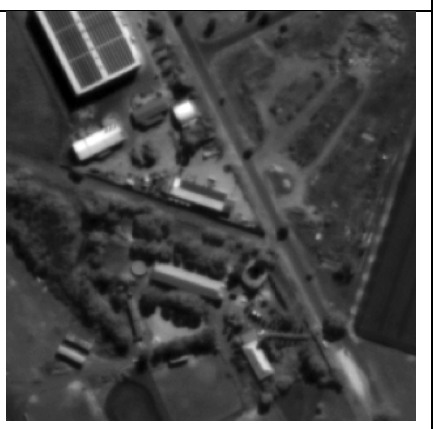
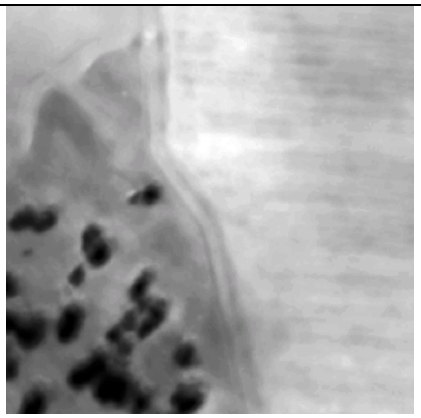
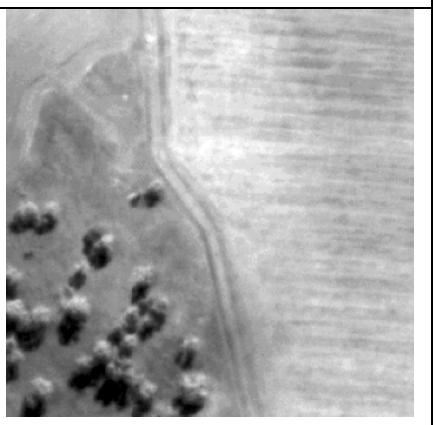
In this work, we selected 102 non-repeat less cloudy or cloud-free Deimos-2 EO images for the GAN network training. The 102 Deimos-2 MS green band images (at 4m resolution) formed 18,782 (256 by 256 pixels) LR training samples. The corresponding 102 Deimos-2 PAN band images (1m resolution) formed 18,782 (1024 by 1024 pixels) HR training samples. Multi-angle Deimos-2 MS green band images for selected sites at Dubai and Adelaide are used for SRR processing and comparison to the higher resolution PAN band images. In this example, a total of 8 Deimos-2 MS green band images [Table 1] are used as inputs. We then crop the roughly aligned input images to smaller samples (size 2048 by 2048 pixels) for the MSA-FAST-CNN-GPT stage processing. The resulting output SRR images (size 4096 by 4096) are then further divided into smaller sample sizes (128 by 128 pixels) to be used for GAN refinement, resulting in the final SRR images (size 256 by 256) with an up-scaling factor of 4.

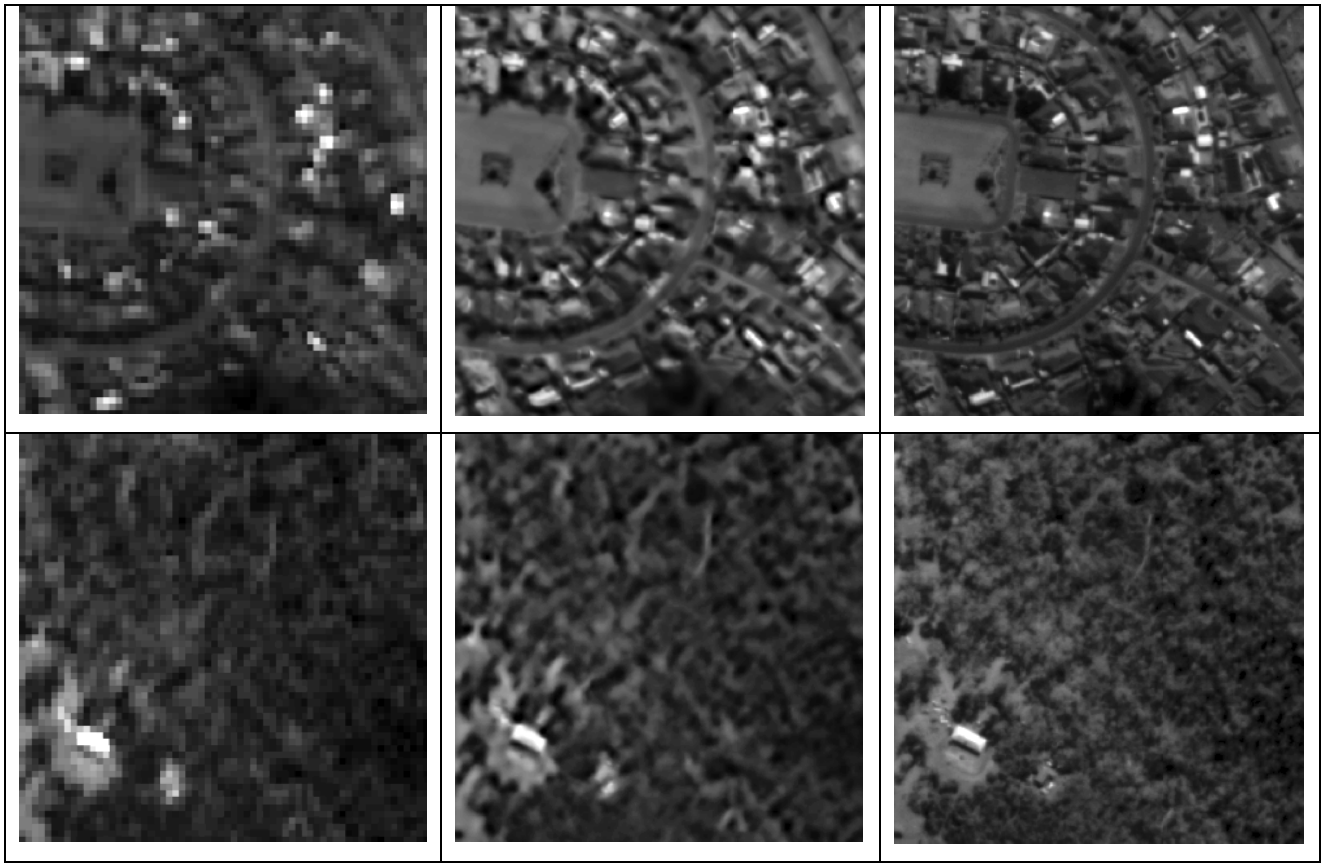
*Table 1. List of the 8 input Deimos-2 MS green band images*

| Image ID | Incidence angle | Date | Data type |
|---|---|---|---|
| DE2_MS4_L1B_000000_20180311T021248_20180311T021253_DE2_20183_DE02 | 6.2º | 2018-03-11 | L1B |
| DE2_MS4_L1B_000000_20180323T020548_20180323T020551_DE2_20361_3270 | 20.0º | 2018-03-23 | L1B |
| DE2_MS4_L1B_000000_20180324T022115_20180324T022119_DE2_20376_DE02 | -10.8º | 2018-03-24 | L1B |
| DE2_MS4_L1B_000000_20180405T021423_20180405T021427_DE2_20554_DE02 | 3.4º | 2018-04-05 | L1B |
| DE2_MS4_L1B_000000_20180411T021057_20180411T021102_DE2_20643_DE02 | 10.4º | 2018- | L1B |

| | | | |
|---|---|---|---|
| | | 04-11 | |
| DE2_MS4_L1B_000000_20180417T020733_20180417T020738_DE2_20732_DE02 | 17.0º | 2018-04-17 | L1B |
| DE2_MS4_L1B_000000_20180506T021306_20180506T021310_DE2_21014_DE02 | 6.3º | 2018-05-06 | L1B |
| DE2_MS4_L1B_000000_20180430T021622_20180430T021627_DE2_20925_DE02 | -0.5º (reference) | 2018-04-30 | L1B |

*Table 2 Example 1m MAGiGAN SRR results in comparison with the original 4m MS green band (left column) and 1m PAN band Deimos-2 images (right column)*

| 4m Deimos-2 MS green band | 1m MAGiGAN SRR | 1m Deimos-2 PAN band |
|---|---|---|
|  |  |  |
|  |  |  |



[Table 2] shows examples of the original 4m Deimos-2 MS green band images (left) in comparison to the MAGiGAN SRR results (middle) and 1m Deimos-2 PAN band images (right).

## 4. S UMMARY AND FUTURE WORK

We have described a SRR system based on the GPT-SRR and deep learning algorithms, called MAGiGAN, applicable to multi-angle EO imaging datasets providing a factor of 4x enhancement in resolution. The MAGiGAN system not only retrieves subpixel information from multi-angle distorted features, but also uses the GAN network to retrieve high-frequency texture details. The MAGiGAN system has been integrated onto a space-qualified NVIDIA® Jetson TX-2 GPU board for speeded-up processing. It has been tested with multiple EO instruments and will be tested with Mars images in the near future. Experimental results in this paper are produced using the UrtheCast Corp. Deimos-2 images.

## ACKNOWLEDGEMENTS

The research leading to these results has received funding from the UK Space Agency Centre for Earth Observation Instrumentation (UKSA-CEOI-10 2017-2018) under SuperRes-EO project agreement n° RP10G0435A05. The authors would like to thank C. Rampersad and colleagues at UrtheCast Corp. for providing the Deimos-2 images.

## REFERENCES


[1] Tao, Y. and J.-P. Muller, "A novel method for surface exploration: super-resolution restoration of Mars repeat-pass orbital imagery," Planetary and Space Science, 121, pp.103-114 (2016).

[2] Shin, D., and J.-P. Muller, "Progressively weighted affine adaptive correlation matching for quasi-dense 3D reconstruction", Pattern Recognition, 45(10): p. 3795 -3809 (2012).

[3] Tao, Y. and J.-P. Muller, "Quantitative assessment of a novel super-resolution restoration technique using HiRISE with Navcam images: how much resolution enhancement is possible from repeat-pass observations", ISPRS 2016 archive, Commission IV/8 (2016).

[4] Bridges, J.C., J. Clemmet, D. Pullan, M. Croon, M.R. Sims, J-P. Muller, Y. Tao, S.-T. Xiong, A. D. Putri, T. Parker, S.M.R. Turner, J.M. Pillinger, "Identification of the Beagle 2 Lander on Mars", Royal Society Open Science, 4(10), p.170785 (2017).

[5] Tao, Y., J.-P. Muller, and W. D. Poole, "Automated localisation of Mars rovers using co-registered HiRISE-CTX-HRSC orthorectified images and DTMs", Icarus, 280, pp.139-157 (2016).

[6] Rosten E. and T. Drummond, "Machine Learning for High-Speed Corner Detection", In: Leonardis A., Bischof H., Pinz A. (eds) Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science, vol 3951. Springer, Berlin, Heidelberg (2006).

[7] Fischer, P., A. Dosovitskiy, T. Brox, "Descriptor matching with convolutional neural networks: a comparison to sift", arXiv preprint arXiv:1405.5769 (2014).

[8] Guo, R., Q. Dai, D. Hoiem, "Single-image shadow detection and removal using paired regions", In Computer Vision and Pattern Recognition (CVPR), IEEE Conference on (pp. 2033-2040). IEEE (2011).

[9] Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", In CVPR, Vol. 2, No. 3, p. 4 (2017).

[10] Rother, C., V. Kolmogorov, A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts", In ACM SIGGRAPH 2004 Papers (SIGGRAPH '04), Joe Marks (Ed.). ACM, New York, NY, USA, 309-314 (2004).

[11] Chang, C.-C., and C.-J. Lin, "LIBSVM: a library for support vector machines", ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27 (2011).

[12] Dosovitskiy, A., P. Fischer, J.T. Springenberg, M. Riedmiller, T. Brox, "Discriminative unsupervised feature learning with exemplar convolutional neural networks", IEEE transactions on pattern analysis and machine intelligence, 38(9), pp.1734-1747 (2016).

[13] Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial nets", In Advances in neural information processing systems, pp. 2672-2680 (2014).

[14] Arjovsky, M. and L. Bottou, "Towards principled methods for training generative adversarial networks", arXiv preprint arXiv:1701.04862 (2017).

[15] Arjovsky, M., S. Chintala, L. Bottou, "Wasserstein GAN", arXiv preprint arXiv:1701.07875 (2017).