



**HAL**  
open science

# An Advanced Benchmarking for Image Compositing Evaluation

Roberto Giudici, Luc Courtrai, Sébastien Lefèvre

► **To cite this version:**

Roberto Giudici, Luc Courtrai, Sébastien Lefèvre. An Advanced Benchmarking for Image Compositing Evaluation. ISPRS International Workshop on Unmanned Aerial Vehicles in Geomatics (UAV-g), 2019, Enschede, Netherlands. pp.303-308, 10.5194/isprs-archives-XLII-2-W13-303-2019 . hal-02343897

**HAL Id: hal-02343897**

**<https://hal.science/hal-02343897>**

Submitted on 13 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# AN ADVANCED BENCHMARKING FOR IMAGE COMPOSITING EVALUATION

Roberto Giudici\*, Luc Courtrai, Sebastien Lefevre

Univ. Bretagne Sud – UMR 6074 IRISA, F-56000 Vannes, France  
giudici.e1506045@etud.univ-ubs.fr, luc.courtrai@irisa.fr, sebastien.lefevre@irisa.fr

**KEY WORDS:** Mosaicing, Virtual Camera, quality assessment, structural fidelity

## ABSTRACT:

This paper introduces a novel benchmarking tool for measuring the robustness of existing mosaicing algorithms in presence of a given set of disturbances. The process combines a set of partially overlapping images into a wide-view result used to represent UAV image series and orthophotography. Geometrical misalignments caused by perspective error may lead to some unpredictable artifacts and phantom effects in the mosaics. A very few solutions measure their immunity to known distortions and mainly focus on registration accuracy measurement. Only limited attention was given to characterize the response of the actual image fusion algorithms and their capacity to properly preserve content geometries. In this paper, we also introduce a new fidelity metric assessing the mosaicing response to a disturbance of a given extent used as prior information. The metric helps to better define the use cases fulfilling aerial imaging requirements.

## 1. INTRODUCTION

Mosaicing refers to the complete process of merging seamlessly a set of adjacent narrow-view images into a wide-view result, providing a significantly larger FOV. The mosaicing pipeline requires geometrical registration and warping prior to any image merge application. A strategy must be adopted when merging sequences are shot at different times to ensure object motion integrity and avoid visual double contours or phantom artifacts to show up. Satellite and high-altitude images appear as static and are mostly immune to movement. UAV low-altitude flights can conversely be severely impacted by motion leading to structural inconsistencies. Existing benchmarks measure the likelihood of images cropped through a Virtual Camera (VC) and are mostly incapable of assessing any quality loss issued from photometric, geometrical, camera noise or vibration disturbances.

The main contribution of this paper relies in the definition of a novel objective assessment coupling (Disturbance, Mosaicing) to compare the resiliency of existing algorithms. The fidelity measures the quality loss after application of the disturbance function to a nominal (ground truth) image through structural PSNR and perceptual SSIM likelihood index, according to their mosaicing response.

Section 2 introduces the main mosaicing principles and methods, before we review existing benchmarking solutions in Sec. 3. Section 4 describes our contribution along with image distortions and assessment metrics. It is used to conduct an experimental analysis in Sec. 5, before we conclude the paper.

## 2. IMAGE MOSAICING: PRINCIPLES AND METHODS

A mosaicing pipeline is mainly composed by two steps: registration and composition (see Fig. 1). The first step aligns a set of images with a reference one, correcting perspective disparities that may cause parallax errors as described by (Adel, 2014). SIFT (Lowe, 2004) and SURF (Bay et al., 2008) are the

\*Also with LGO UMR 6538 CNRS UBO UBS

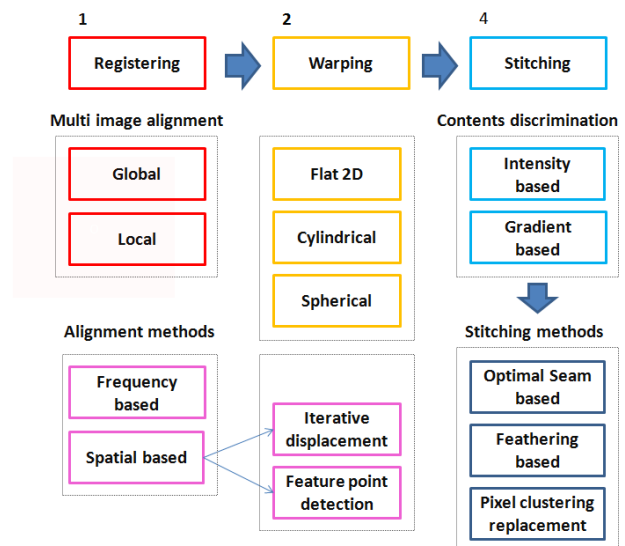


Figure 1. Mosaicing pipeline.

most robust feature point extraction algorithms to scaling and affine distortions. The FAST (Rosten, 2018) corner detection is a rapid but not scale-invariant solution used in real-time applications.

(Sheikh Faridul et al., 2014) propose a novel method for illumination and device invariant stitching capable of withstanding and correcting large changes in illumination. (Ghosh, Kaabouch, 2016) and (Xu et al., 2016) propose methods for correcting perspective distortions in UAV imaging.

Among existing methods, we can mention:

- Image composition: Blending and best-seam stitching are the most common techniques used to merge different contents, as proposed by (Botterill et al., 2010);
- Best-Seam search: The seam is placed between two images where intensities are minimal in the area of overlap.

- Gradient based approach: This class of algorithms searches the seam line featuring the lowest photometric difference between overlapping regions of adjacent images. (Levin et al., 2004) propose to perform seamless image stitching in the gradient domain (GIST) to minimize a cost function based on the dissimilarity to each of the input images and the visibility of the resulting seam.
- Watershed: This transform proposed by (Soille, 2006) and (Worringer et al., 2003) is a morphological marker-controlled segmentation.
- Graph-cut optimization: In this transform, a label associates each pixel to an image. The basic technique is a labeling problem solved through an energy minimization function graph, e.g. maximum flow or MinCut algorithm. (Dijkstra, 1959) and (Bellman, 2010) are two historically known algorithms for graph shortest path search. (Uyttendaele et al., 2001) propose a block-by-block based algorithm based on Regions of Difference (ROD), representing areas of overlap between input images. (Agarwala et al., 2004) propose a graph-cut supervised multi-label optimization to compute the weight of a region in pixel-wise mode. An iterative alpha-expansion graph-cut algorithm is used. (Brown, Lowe, 2007) formulate stitching as a multi-image matching invariant local features. (Gracias et al., 2009) divide the mosaic space into large disjoint Regions of Image Intersection (ROI) with a camera-to-center penalization function.

### 3. RELATED WORK

(Bevilacqua, Azzari, 2006), (Boutellier et al., 2008) and (Weibo et al., 2013) propose Virtual Camera (VC) and Object Indicator based plans for registration accuracy assessment in image sequences. (Li-hui Zou, 2011) propose a pair-wise blind mosaicing framework with use case recommendation matrices. (Achanta et al., 2012) is an example of a subjective metric based on Human Visual System (HVS) response analysis. (Khan et al., 2012) propose a new quantitative metric for assessing the quality of mosaic to the input narrow images. The process computes the Structural Similarity Index (SSIM) of the high-frequency information (HFI) for structural assessment and the Spectral Angle Mapper (SAM) computed on low-frequency for photometric assessment. The existing solutions measure registration accuracy by relying on well-established PSNR and perceptive SSIM metrics and may not be suitable for any structural or geometrical consistency assessment.

(Bevilacqua, Azzari, 2006) and (Boutellier et al., 2008) focus on cumulative error measurement in a mosaicing data set. The benchmark uses real-world remote sensing scenes. Satellite, high and low-altitude UAV images taken under different weather and lightening conditions have been chosen. Workflow of the proposed VC framework is given in Fig. 2.

### 4. PROPOSED BENCHMARKING

We propose a novel method for direct evaluation of a distortion / mosaicing algorithm with fully aligned images (planar constraint). It measures the response of a set of algorithms under probe to a set of UAV typical operational disturbances, producing an a priori index of robustness. The image composition is computed through direct application of blending

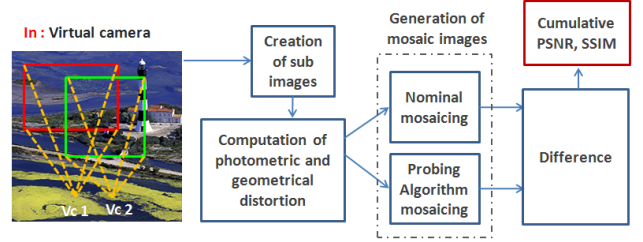


Figure 2. Virtual camera benchmarking

and optimal-seam stitching methods applied on a pair of input narrow-view icon VC crops. The assessment requires first to define a formal metric to measure the robustness of the algorithms in presence of disturbances.

#### 4.1 Mosaicing comparative assessment

We now introduce a mathematical formalization of our benchmark and detail the resulting pipeline. We consider a set of test cases  $T = \{t_0, \dots, t_{nt-1}\}$ , where  $nt$  is the number of test cases. For each test case, we assume a set  $S_t$  of 24-bit RGB input images of size  $nm$  pixels, with  $S = \{x_0, \dots, x_{ns-1}\}$  and  $ns$  the size of the imageset. A pair of consecutive images with  $x_{k-1}$  and  $x_k$  the first and second images respectively, with  $k \in [0, n-1]$ .  $\Omega = \{\omega_0, \dots, \omega_{nm-1}\}$  is the set of mosaic images and the set of mosaic operators is written  $M = \{m_0, \dots, m_{nm-1}\}$ , with  $i \in [0, nm-1]$ . We also have  $m_i \in M$ , where  $\omega_i = m_i(x_{k-1}, x_k)$  is the mosaic image issued from the application of the operator  $m_i$ , and  $nm$  denotes the number of defined mosaicing functions. We also define a reference mosaicing operator  $m_{ref} = x_{k-1} \cap x_k$  based on the extraction of the overlapping region. A set of distortions  $D = \{d_0, \dots, d_{nd-1}\}$ , with  $x'_k = d_j(x_k)$  is finally defined. The distortion  $d_j$  is applied to the image  $x_k$ , with  $j \in [0, nd-1]$  and where  $nd$  is the number of distortions defined in the current workbench.

We compute the reference by image content replacement through  $\omega_{ref} = m_{ref}(x_{k-1}, x'_k)$  and the probe mosaic with the algorithm under probe using  $\omega_{probe} = m_i(x_{k-1}, x'_k)$ . We measure the final quality of a stitch between two different contents with no prior information available as the likelihood between  $\omega_{ref}$  and  $\omega_{probe}$ , considering their region of overlap  $S_\omega = x_{k-1} \cap x'_k$ . Supervised mosaicing may take advantage of prior information such as relevance or age of the image in a sequence, to gather higher rank and thus larger surfaces on the final mosaic. When two distinct images with the same relevance are merged, the best result would ideally be composed of 50% of each image to minimize loss of contents. At a first glance, we may think that if we generate an overlapping region issued of the undistorted  $x_{k-1}$  image, the  $S_{\omega_{ref}}$  will be totally error-free, and the assessment will penalize any distortion introduced by the  $S_{\omega_{probe}}$ . As such, the best solution would ideally contain no potentially distorted contents issued from the  $x'_k$  image in the resulting mosaic, leading to image losses. We decided to add the  $S_{\omega_{ref}}$  case computed with the fully distorted image  $x'_k$  and average the results to improve reliability. Our approach computes a PSNR and SSIM measuring the difference of contents in shared regions generated by  $S_{\omega_{ref}}$  and  $S_{\omega_{probe}}$ :

$$PSNR_{ref} = \frac{1}{2} \left( PSNR(S_{\omega_{ref1}}, S_{\omega_{probe}}) + PSNR(S_{\omega_{ref2}}, S_{\omega_{probe}}) \right) \quad (1)$$

$$SSIM_{ref} = \frac{1}{2} \left( SSIM(S_{\omega_{ref1}}, S_{\omega_{probe}}) + SSIM(S_{\omega_{ref2}}, S_{\omega_{probe}}) \right) \quad (2)$$

The principle is summarized in Fig. 3. The benchmark applies increasing levels of disturbance to the image  $x_k$ . The  $S_{\omega_{probe}}$  may change according to the robustness of the algorithm under probe subject to the selected disturbance. The metrics are computed for every different couple of  $(d_j, m_i)$  and summed-up for all the imageset images. Single mosaicing, single distortion PSNR is computed on the nominal input  $x_k$  and distorted  $x'_k$ , giving the  $PSNR_{i,j,k}$  and  $SSIM_{i,j,k}$ . The values of  $PSNR_{i,j,k}$  and for the  $SSIM_{i,j,k}$  are then summed up image by image for entire the imageset:

$$PSNR_{i,j} = \sum_{k=0}^{nk-1} PSNR_{i,j,k}^{ref}, \quad (3)$$

$$SSIM_{i,j} = \sum_{k=0}^{nk-1} SSIM_{i,j,k}^{ref}. \quad (4)$$

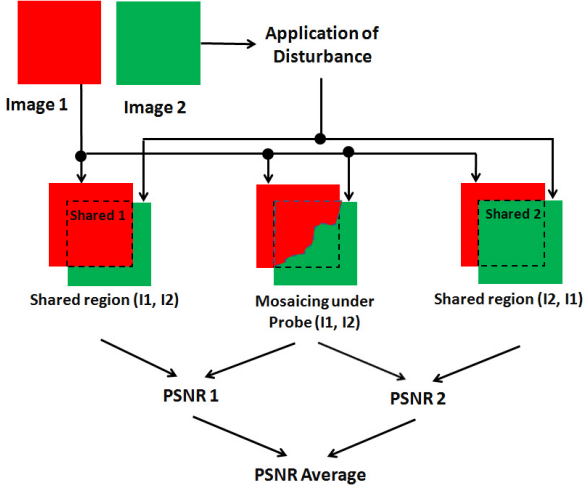


Figure 3. Process for mosaicing assessment

## 4.2 Benchmarking Pipeline

The first step is to apply VC. To do so we generate ROIs (Regions of Interest), and crop  $x_{k-1}$  and  $x_k$  with a  $(x_0, y_0)$  fixed shift. We then apply a disturbance function to the current image  $x_k$  to gather the image  $x'_k$ , before computing a nominal mosaicing  $\omega_k = m_i(x_{k-1}, x'_k)$  by overlapping the ROI  $x'_k$  over  $x_{k-1}$ . We then compute the  $PSNR_{i,j}$ ,  $SSIM_{i,j}$  quality assessment metrics for each imageset. Finally, we aggregate these PSNR and SSIM measures to build assessment reports and compute the  $MSE_{PSNR_j}$  and  $MSE_{SSIM_j}$ .

The mosaicing fidelity metric assesses the likelihood variance at distortion time as prior information to the mosaicing likelihood, measured through the PSNR and SSIM on the entire set of results. (Khan et al., 2009) demonstrate that the pixel-wise PSNR is not meaningful to measure photometric and geometrical distortions occurring at the same time. Conversely, our index is composed of a cost function penalizing strong discordance through PSNR and SSIM likelihood measured after application of a single distortion.

A low distortion variance generating high variance at mosaicing will be highly penalized as it may lead to unpredictable results. High distortion variance and low mosaicing variance corresponds to a reduced algorithm sensitivity. The fidelity workbench pipeline is illustrated in Fig. 4. We define  $\bar{P}_d = (PSNR_d)^\gamma$  the normalized value of the linearized distortion PSNR, with  $\gamma = 0.2$  an empirically-defined linearization coefficient. We also note  $\bar{P}_m = PSNR_m$  the normalized value of the mosaicing PSNR. We then compute the PSNR fidelity  $F_{psnr} = (P_d - P_m^\lambda)^2$ . We assign a higher weight to the mosaicing  $P_m$  than  $P_d$  and we set the weight coefficient  $\lambda = 0.5$  assigned to the mosaicing operator. SSIM fidelity can be computed following a similar reasoning.

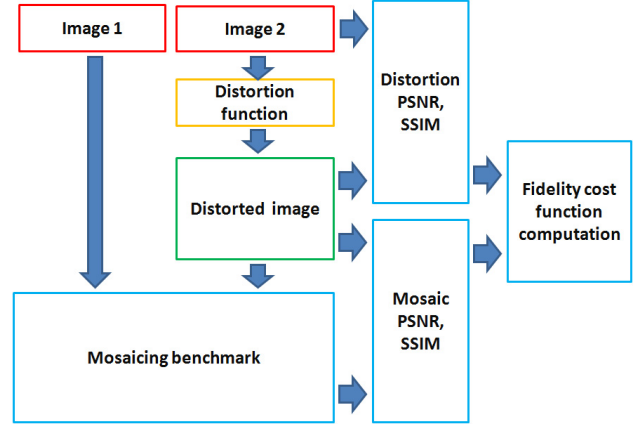


Figure 4. Fidelity computation workflow

## 5. EXPERIMENTS

### 5.1 Scenarios and parameter settings

We run our benchmark with the following mosaicing algorithms: feathering, multiband (blending); Watershed, Graph-cut, Voronoi (stitching).

As far as the dataset is concerned, we consider various remote sensing and UAV data sources. More precisely, each imageset is made of 20 images coming from: SPOT 5, 6, 7 and Pleiades satellite imagery (satellite), coastal UAV surveys (coastal), black and white historical photographs (BW), various images from the geomatics lag of Rennes 2 University (Costel), and from the ISPRS 2D semantic labeling contest (ISPRS).

We then define 6 test cases to conduct experimental comparisons:

1. Change in image overlap: with the coastline imageset, we measure the robustness when the overlap ratio of the input images changes.
2. Change in number of distortion steps: with the coastline imageset, we measure the robustness when the number of distortion steps changes.
3. Change in number of input images: with the coastline imageset, we measure the robustness when the imageset size changes.
4. Change in imageset: we consider BW, Coastline, and Satellite imagesets and measure the robustness when different imagesets of the same size are considered.

5. Reference ISPRS imageset: we measure the fidelity and robustness to a set of ISPRS images that are well-established for benchmarking activities.
6. Change in ISPRS number of images: we measure the robustness to growth in the image size.

Table 1 lists the parameters applied to each testing scenario.

## 5.2 Results

Figure 5 groups the different responses by distortion. The photometric distortions induce the highest levels of PSNR variability. Hue and Vignetting give the worst photometric response. All test cases are robust to geometrical linear, affine distortions and noise (with the exception of the thermal noise and Gaussian blur). Test 4 (Black and White imageset) yields to the lowest variability.

Figure 6 shows the perceptual geometrical and noise acceptable variabilities. 9) shows the good performance of watershed has the highest. Graph-Cut and Voronoi are observed to lack of robustness to photometric changes (see Fig. 10). Multiband gives far better structural results than feathering, as expected. Figure 11) show blending high perceptual invariance to noise and a low invariance to geometrical transforms. Figure 7) confirms how the photometric distortions harm severely the structural variability of the Graph-Cut and Voronoi. Figure 8) confirms how the geometrical transforms harm the perceptual variability of the Graph-Cut and Voronoi, with acceptable results on blending. Figures 12 and 13 show how watershed produces the highest structural and lowest perceptual robustness on the same benchmarking scenario and is robust to photometric changes. Feathering has the best computational speed (597.337 sec.) and is widely used in real-time applications, but may not preserve structural integrity. Optimal-seam watershed, Graph-Cut and Voronoi may sacrifice the visual pleasantness to preserve the structural integrity required by geomatics applications. Watershed is the best trade-off between pleasantness and structural preservation with an acceptable computational time (893.181 sec.). Graph-Cut shows the worse computational performances (18285.4 sec.).

We have introduced the fidelity analysis metric to measure the dependency between image distortion and mosaicing, as well to penalize high MSE discordances on the entire imageset. It has been tested on the imageset 4 featuring the following subsets: Satellite, Coastline and BW. The Satellite subset features the highest homogeneity with images with similar subjects. Figure 14 shows how the geometrical disturbances give the highest structural fidelity. The heterogeneous Coastline imageset contents result in lowest structural fidelity caused by unpredictable mosaicing results. The BW imageset has the highest structural and perceptual fidelity. Figure 15 shows how geometrical distortions produce the lowest perceptual fidelity. Photometric distortions seem also to highly penalize perceptual fidelity. A set of recommendations and use cases are resumed in Table 2.

## 6. CONCLUSION

In this paper, we have introduced a novel benchmarking tool aimed to classify state-of-the-art UAV and remote sensing mosaicing algorithms upon structural and perceptual robustness in presence of a set of real-life operational disturbances with

different extents. The results of a set of exhaustive tests show how the watershed-based mosaicing features far better structural invariance to the entire set of disturbances. Feathering is perceptually more robust when it is subject to the same disturbances. Geometrical distortions give a high structural but low perceptual fidelity. None of the algorithms appears to be robust to noise, large photometric distortions and vignetting. Monochrome leads to the highest fidelity and should preferably be used to get the best mosaics. Despite the features introduced in this paper, we believe the proposed work needs to be pursued to fulfill the needs of aerial imaging professionals. A finer benchmark should then be created to measure the variance and fidelity of a set of simultaneous disturbances occurring on a single image. Ideally, the tool should be capable of assessing an entire sequence of images including aircraft altitude and position metadata vectors, featuring finer analysis and providing far more reliable results for each different flight. Nevertheless, the distortion ranges have been chosen to fit real-world operational conditions.

## REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, Sabine, 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 2274–2282.
- Adel, E., 2014. Image Stitching based on Feature Extraction Techniques: A Survey. *International Journal of Computer Applications*, 99, 975–8887.
- Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M., 2004. Interactive digital photomontage. *ACM SIGGRAPH*.
- Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110, 346–359.
- Bellman, Richard, 2010. *Dynamic programming*. Princeton University Press.
- Bevilacqua, A., Azzari, P., 2006. High-Quality Real Time Motion Detection Using PTZ Cameras. *AVSS*.
- Botterill, T., Mills, S., Green, R., 2010. Real-time aerial image mosaicing. *ICVNZ*.
- Boutellier, J., Silvén, O., Tico, M., Korhonen, L., 2008. Objective evaluation of image mosaics. *International Conference on Computer Vision and Computer Graphics*, 107–117.
- Brown, M., Lowe, D.G., 2007. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*, 74, 59–73.
- Dijkstra, E.W., 1959. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1, 269–271.
- Ghosh, D., Kaabouch, N., 2016. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34, 1–11.
- Gracias, N., Mahoor, M., Negahdaripour, S., Gleason, A., 2009. Fast image blending using watersheds and graph cuts. *Image and Vision Computing*, 27, 597–607.

Benchmark type	Imageset	Image size	Nbr. Img	Nbr. Dist.	Dist. strength	Overlap ratio
Variable overlap ratio	Coastline	2000	1	1	4/10	25%, 50%, 75%
Variable dist. step	Coastline	2000	1	2	4/10, 8/10	25
Nbr. images	Coastline	2000	2, 4, 8	1	4/10	25
Nbr. subjects	All	2000	20	1	4/10	25
Reference	ISPRS	2000	20	1	4/10	25
Reference size image	ISPRS	200, 500, 1000	20	1	4/10	25

Table 1. Test cases

Algorithm	Geomatics	Photography and media	Structural preservation	Real-Time	Computational time
Feathering	No	No	No	Yes	Fastest
Multiband	No	No	No	Yes	Fast
Watershed	Yes	Yes	Yes	No	Slow
Graph-Cut	Yes	Yes	Yes	No	Slowest
Voronoi	Yes	Yes	Not good	Yes	Fast

Table 2. Table of mosaicing use cases

Khan, M.M., Alparone, L., Chanussot, J., 2009. Pansharpening Quality Assessment Using the Modulation Transfer Functions of Instruments. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3880–3891.

Khan, M.M., Hafiz, R., Cho, Y., Qureshi, H.S., Cha, J., 2012. Quantitative quality assessment of stitched panoramic images. *IET Image Processing*, 6, 1348–1358.

Levin, A., Zomet, A., Peleg, S., Weiss, Y., 2004. Seamless image stitching in the gradient domain. *ECCV*.

Li-hui Zou, Jie Chen, Juan Zhang, 2011. Assessment approach for image mosaicing algorithms. *Optical Engineering*, 37–41.

Lowe, David G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 91–110.

Rosten, E., 2018. FAST corner detection.

Sheikh Faridul, H., Stauder, J., Treméau, A., 2014. Illumination and device invariant image stitching. *ICIP*.

Soille, P., 2006. Morphological image compositing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 673–683.

Uyttendaele, M., Eden, A., Szeliski, R., 2001. Eliminating Ghosting and Exposure Artifacts in Image Mosaics. *CVPR*.

Weibo, Z., Jianxun, L., Zhi, Z., 2013. Performance evaluation approach for image mosaicing algorithm. *CCDC*, 3786–3791.

Worring, M., Van den Boomgaard, R., Guyen, H., 2003. Watersnakes: Energy-driven watershed segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 330–342.

Xu, Y., Ou, J., He, H., Zhang, X., Mills, J., 2016. Mosaicking of Unmanned Aerial Vehicle Imagery in the Absence of Camera Poses. *Remote Sensing*, 8, 204.

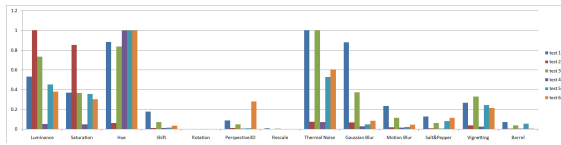


Figure 5. Distortion - Test Normalized PSNR MSE Response

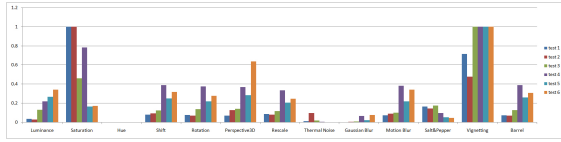


Figure 6. Distortion - Test Normalized SSIM MSE Response

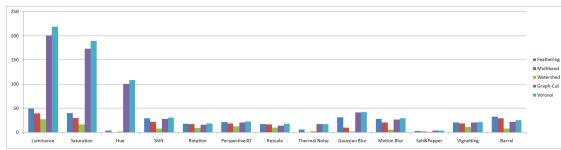


Figure 7. Distortion - Mosaicing PSNR MSE Response

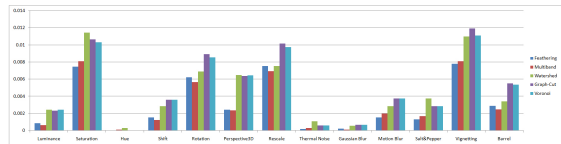


Figure 8. Distortion - Mosaicing SSIM MSE Response

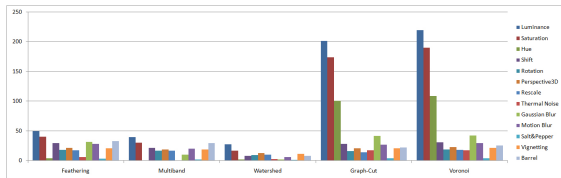


Figure 9. Mosaicing - Distortion PSNR MSE Response

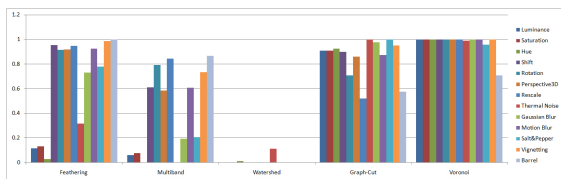


Figure 10. Mosaicing - Distortion PSNR MSE Response Normalized

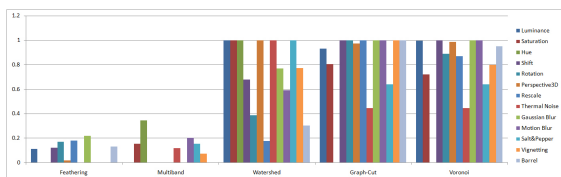


Figure 11. Distortion - Mosaicing SSIM MSE Response Normalized

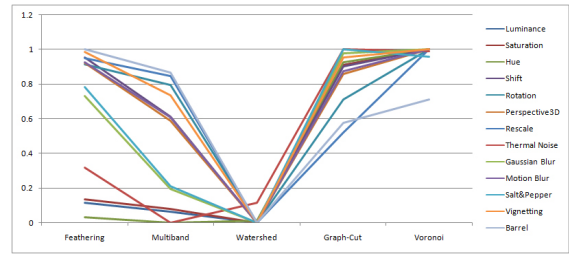


Figure 12. Mosaicing - Distortion PSNR MSE Normalized Response

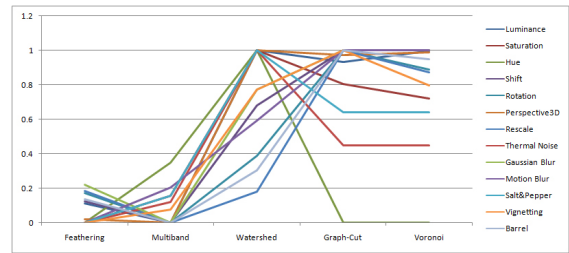


Figure 13. Mosaicing - Distortion SSIM MSE Normalized Response

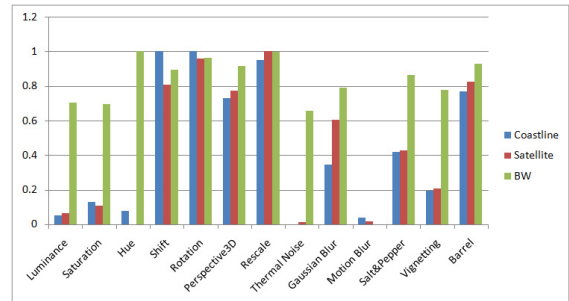


Figure 14. Distortion - Mosaicing PSNR Fidelity

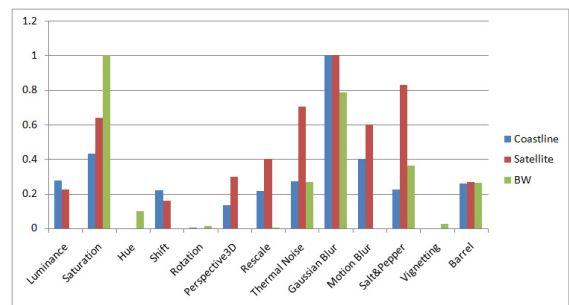


Figure 15. Distortion - Mosaicing SSIM Fidelity