

MOSAICING FIDELITY GEOMETRICAL ASSESSMENT BASED ON SURF POINT CLASSIFICATION

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, Geometrical Fidelity Assessment, SURF

ABSTRACT:

The recent development of automated UAV imaging applications for geomatics is leading to an unprecedented rapid growth in mosaic and DEM images. Newer, more advanced algorithms are being extensively studied to fulfill constantly increasing requirements. Sequences of georeferenced images that cannot be shot at once are merged with ideally no geometrical distortion to allow for 3D reconstruction and orthophotography generation. If newer robust mosaicing algorithms are being developed to withstand UAV tighter real-time constraints, very little attention has been given to objectively assess their capacity to coherently combine different image geometries. Changes in FOV and illumination, oblique scenes and dynamic sequences featuring movement and object occlusion are not yet fully handled by state-of-the-art algorithms. Despite the large panel of existing applications featuring different quality constraints, our work focuses on the geomatics context, where the main requirement is the respect of geometrical proportions among images. Geometrically coherent mosaics and 3D reconstructions can only result from fully static, high-altitude orthogonal views. GCPs or well known geographical landmarks are employed with RTK-GPS for scene georeferencing. The overall fidelity is computed as the global variance of the distances between SGPs. However, very few ways seem to exist to measure the content coherence of large geographical scenes, interiors or unreferenced scenes. Distances are mostly expressed in a local coordinate system and GCPs are often unavailable. The main contribution of this paper consists in the definition of a formal metric for measuring local coordinate geometrical fidelity of mosaics issued of UAV image sequences, in real operating conditions. A Mosaicing Fidelity Assessment (MSF) index is therefore computed for every couple of consecutive images of a sequence. The metric generates an index based on the distances of SURF feature points extracted in the images and compares them to estimate geometrical changes transferred to the mosaic. The solution can detect perspective inaccuracies caused by residual registration errors. It can fulfill the requirements of most of imaging and geomatics applications and can be executed as nearly real-time due to its low computational complexity.

1. INTRODUCTION

The development of automated vision and 3D acquisition systems in remote sensing and geomatics requires to compose different visual contents at different times and different perspectives into a single wide-view image. Besides pure photographic usage, research is currently focusing on algorithms capable of respecting the geometry of a scene represented by the images being stitched. Basic feathering algorithms are being replaced by more advanced stitching solutions capable of respecting the natural boundary of real objects contained in the scenes. Nevertheless these algorithms are blind and not capable of properly characterizing the objects contained in a scene and preserving their geometry and integrity while mosaicing image sequences. Handling dynamic objects with speeds sensibly higher than the frame rate and under highly changing illumination conditions can hardly be achieved by existing algorithms and may result in artifacts and phantom effects in the regions of boundary overlap. As mosaicing algorithms are not capable of preserving the image geometry, we need to measure *a posteriori* the resulting structural coherence, i.e. after the mosaic generation. Since geo-referencing must be recomputed after mosaicing, the usage of GCPs (Ground Control Points), or other static objects with known shape and position, is normally exploited for geo-referencing the entire mosaic. Mean distance and variance between GCPs gives the overall accuracy of the geo-referencing process. Unfortunately, the GCP points can not always be deployed in interior geo-unreferenced

scenes or large natural, hardly accessible sites. Furthermore, mosaicing may involve entities expressed in local coordinates, where geo-mapping is not required.

All the aforementioned cases may need the development of a further geometrical accuracy metrics capable of measuring the extent of perspective and morphological changes across time independently of existing GCPs. Pure image based mosaic geometrical fidelity measurement would give its reliability and detect images with insufficient precision. A posteriori measurement may be also used as prior to better drive existing algorithms to enhance geometrical boundary respect. A mosaic is mainly built from the composition of two (or more) semi-overlapping images, with an overlap ratio of about 70% in aerial imagery. The contents of the first image are stitched with the second image in their overlapping region and along a seamless path crossing object boundaries or uniform region, where no texture was detected. If objects belonging to this region move or their shape changes, the algorithm should be able to transfer only the most meaningful and uncorrupted contents to the mosaic.

The algorithm proposed here is a first attempt to objectively measure the extent of different textured contents stitched in the shared region of an image pair. The concept may be easily extended to N-image sequences. The solution compares a pair of input narrow-view images to the generated mosaic and with an ideal undistorted one, that may be represented by a static scene. The geometrical distortions and structural unconformities between two consecutive images are detected by a cost function

* Also with LGO UMR 6538 CNRS UBO UBS

constructed through Euclidean distances measured between SURF points. The algorithm does not simply measure pixel-wise Euclidean distances between images, as the process would be very time consuming and may compute distances on poorly textured or uniform regions, where structural coherence respect is not meaningful. Conversely, distortions on highly textured regions would meaningfully assess coherence loss harming image usability. Our contribution aims to image acquisition manufacturers and UAV geomatic communities, working with badly georeferenced scenes. The algorithm may also be used to assess the response of existing mosaicing algorithms and predict their capacity to preserve object coherence and contribute to enhance their overall robustness.

This paper is organized as follows. After having explained the requirements in image mosaicing to assess structural coherence in Sec. 1, we review the existing methods in Sec. 2 along with the motivation leading to the definition of a new assessment algorithm. We then describe in Sec. 3 the conceptual method, its pipeline and how to embed it into existing mosaicing solutions. Section 4 details the experiments done on some sample imagesets, before concluding the paper in Sec. 5.

2. RELATED WORK

Despite the extensive research conducted to enhance mosaicing robustness, very little attention has been given to the definition of formal metrics to assess algorithm behaviors and results. The large majority of the recent contributions aim in best seam robustness enhancement on blind stitching algorithms. The only objectively measured improvements rely on the capacity of the best seam to better fit photometrical and structural discontinuities that may correspond to object boundaries. No formal metric has ever been defined to measure the structural fidelity of a mosaic. Nevertheless, as the process is executed on fully aligned images, geometrical registration and warping accuracy remain critical to obtain well aligned mosaics.

(Ghosh, Kaabouch, 2016) and (Xu et al., 2016) give an exhaustive overview of the main process along with recent solutions. The authors point out how the accuracy of the entire process depends on the feature point (e.g. SIFT or SURF) generation, which is required to match images shot with different poses and subject to different perspectives. The large majority of the assessment metrics measure the registration error as the main source of structural discontinuities. Best seam stitching will mostly fail while trying to follow poorly aligned or overlapping edges, resulting in double contours and other artifacts. SIFT (Lowe, 2004) and SURF (Bay et al., 2008) are among the most successful perspective invariant feature point algorithms. SURF has been introduced as a computationally improved version of the SIFT based on squared box-filter as approximation of the Gaussian filter to create blob-response maps. It is invariant to affine transforms, rotation, scaling. It is also robust to perspective and illumination. (Li et al., 2016) analyze the effective robustness of the SURF feature points and compare them with other commonly used points.

Mosaicing algorithms have no prior information about the scene contents and can only rely on colour, gradient and texture to compute a seam that does should not cross entity boundaries. Structural inconsistencies and visual artifacts may therefore not be measurable through existing pixelwise image likelihood metrics like EQM, PSNR or SSIM that are mostly used to measure image degradation after filtering, compression and

transmission. (Wang et al., 2004) give an overview of the main metrics used to compare different image contents. (Khan et al., 2012) point how the PSNR would not be reliable when photometric and geometrical distortions apply at the same time, as geometrical distortions occurring at stitching time may be photometrically induced. The SSIM accounts only for perceptual distortions and may not properly distinguish perspective errors if they are hardly visible by human eye. (Boutellier et al., 2008) create an image sequence from the reference image and then compare the generated mosaic with the reference image. (Bevilacqua, Azzari, 2006) propose a mosaicing assessment plan aiming UAV imaging through a virtual camera framework for real-life distortion simulation and apply only a set of pixelwise metrics. The framework from (Ghosh et al., 2012) measures four metrics: percentage of mismatches, difference of pixel intensities, peak signal-to-noise ratio and mutual information between the base (ground truth) image and mosaic. These metrics are computed pixelwise and (Rane et al., 2003) compute the Overall Percentage Closeness (OPC) metric that measures the number of black pixels of the difference image.

A closer look at these frameworks shows how they all measure bitwise differences relating to registration accuracy. Their main drawback is that they would not properly distinguish whether errors occur at registration or stitching time. The Mosaicing Structural Fidelity (MSF) index is introduced in this paper as the difference of structural signatures to overcome limitations of the aforementioned methods. The mean of the extracted SURF point positions (centroids) is used to define the signature of the image geometrical structure, aiming at global image displacement invariance. As the algorithm relies on feature point extraction, its immunity to noise and maximum accuracy will be related to the SURF implicit robustness and may be enhanced as newer, more robust extractors will be available.

3. METHOD

The MSF is computed on differences between SURF point positions between a pair of source images extracted from a time sequence and their resulting mosaic images. Figure 1 shows the distance between two points in image mosaicing and how it is measured by the current algorithm.

The algorithm uses properties of SURF point descriptors to transform geometrical displacements in a given region into some unique positional coordinates. Feature points are extracted in highly textured zones and may not cover uniform regions that are considered irrelevant for any geometrical measurement. SURF points are represented as blobs as they are extracted through an integer approximation of the determinant of a Hessian Blob detector. The actual detection is done by a box filter. The descriptor is composed of the *size* obtained by the number of divisions of the square grid, *response* given by the actual amount of details found in the neighborhood of the point, the *orientation* of the strongest gradients and *number of octaves* required to find the point. The *X, Y position* is also given by the descriptor. Larger sizes will give higher precision, but the process will run slower. (Bay et al., 2008) point out how extended descriptor sizes exhibit better precision, but may yield to poorer recall. Distortions are measured only around the points extracted on highly textured regions. Descriptor entries of a sub-region will have a low response on relatively homogeneous regions. The algorithm does not try to match feature points between images, as severely distorted images may feature completely different points that could hardly be matched. The algorithm first extracts

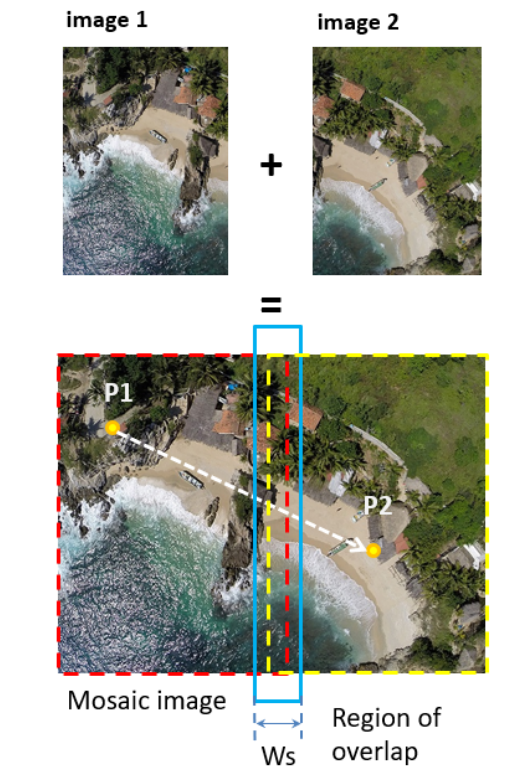


Figure 1. Distance between two points in image mosaicing

the shared region from the left and right original (narrow-view) images and the mosaic. SURF points are then extracted. Figure 2 shows the entire set of points extracted from an image of the imageset under analysis.

Each point is defined by the following signature: (x, y) position, the descriptor response (HoG texture) R , orientation O and size S in pixels. The size tells the maximum distance at which structural and photometric changes will affect the signature of the point. The set of resulting SURF points is sorted by strength and size, where the latter parameter is interpreted here as the maximum reachability distance. From now on, the SURF point size will be intended as reachability and the response will be expressed as sensitivity. Only the K -highest ranked entries are retained and used to compute the Euclidean distance and a centroid. Figure 3 shows the selection and K -ranking process. At each step, a centroid based on a combination of R and S is generated as seen in Figure 4.

Two working modes exist:

- Higher reachability: Finds distortions that are far from the current SURF point, regardless of their strength;
- Higher sensitivity: Finds weaker distortion points, but at within a shorter range

Given σ_i as the generic i^{th} SURF point extracted from the image j , the centroids c_j of the j^{th} image are defined by

$$c_x = \sum_{i=0}^{K-1} x_{\sigma_i}, \quad (1)$$

$$c_y = \sum_{i=0}^{K-1} y_{\sigma_i}. \quad (2)$$

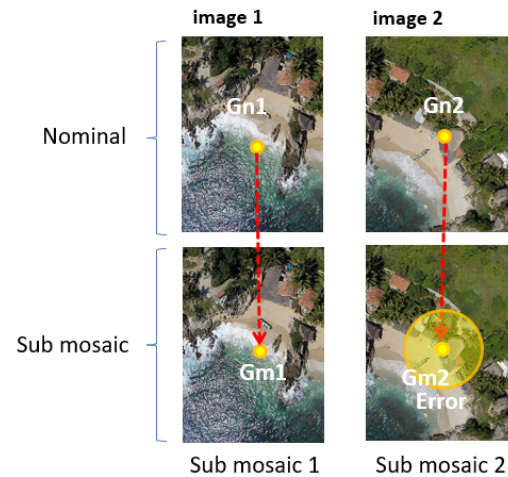


Figure 2. Euclidean distances between centroids

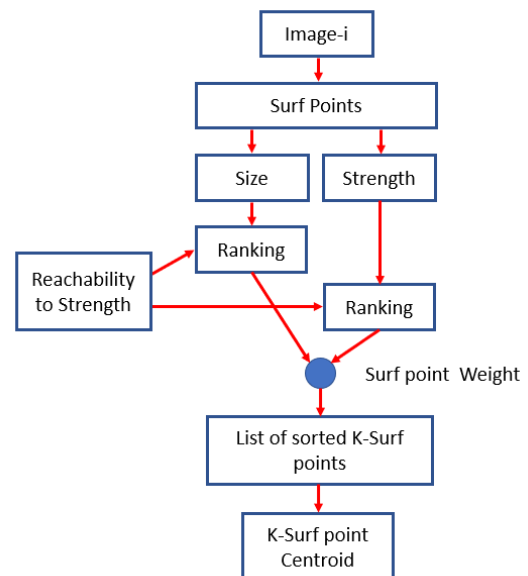


Figure 3. SURF ranking pipeline

where K is the number of highest ranked points. The Euclidean distance of the left and right images is computed through

$$d_{\text{left}} = c_2 - c_0, \quad (3)$$

$$d_{\text{right}} = c_3 - c_1, \quad (4)$$

where the right image is assumed to be the distorted one. Finally, the MSF is defined by

$$MSF = d_{\text{right}} - d_{\text{left}}. \quad (5)$$

A *Sensitivity To Reachability* index is manually provided as input to select a Sensitivity/Reachability ratio that gives the best detection, according to the image type. The minimum ray and strength can be manually selected to gain a better reachability of structural changes (higher ray will find distortions over larger surfaces) despite a lower sensitivity, or higher sensitivity around more restricted points. SURF computed on a sample image produces a large number of unsorted points, with descriptors



Figure 4. K-ranked SURF geometrical centroid



Figure 5. SURF points extracted from a sample image of the imageset under analysis

featuring different sizes and responses as shown in Figure 5. Figure 6 shows the 10% highest response ranked points while Figure 7 shows the 10% highest reachability ranked points.

The highest sensitivity shows how only a restricted area of the image features robust points, mainly extracted on buildings and other artificially crafted objects. The beach and other natural entities would not be properly detected as they would not generate robust points. A good trade-off should therefore be found when working on images featuring artificial and natural entities. Once the centroid has been computed on the shared region of the left and right original (narrow-view) and mosaic images, the centroid of the left image will be subtracted to the centroid computed on the mosaic. The same process is computed between the right image and the mosaic. The process is repeated with a nominal undistorted mosaic, giving the Structural Mosaicing Fidelity of the fully undistorted images. This latter index will be subtracted to the first unbiased MSF, giving the effective unbiased MSF value. The entire pipeline is shown at Figure 8.

Uniform and poorly textured regions around these points are therefore skipped. The maximum number of SURF points used to compute the centroid can be manually defined to improve computational performances at the price of a reduced precision. Even small changes in pixels at a distance included in the SURF ray, will induce a change in the signature and position. Intrinsic SURF robustness to known disturbances fixes its maximum accuracy. The differences of the Euclidean distances computed on the left and right images are then summed up to get the final score that will represent the actual fidelity. Figure 9 shows how SURF points can detect structural and photometric distortions $\{d_1, d_2, d_3\}$ of the points $\{p_1, p_2, p_3\}$, within their ray $\{r_1, r_2, r_3\}$. The distortion point d_1 will be detected by the SURF points p_1 and p_2 , changing its relative signature. d_2 will be detected by p_1 and d_2 will not be detected. The signature of the SURF point p_1 will be influenced by the size (in pixels) and values of the distortions d_1 and d_2 . p_1 is the point with the highest response, suggesting that it was extracted from sharp, relevant points.

4. EXPERIMENTS

The algorithm has been tested on a subset of aerial and UAV images currently used in geomatics. As the algorithm is supposed to measure the structural fidelity of mosaics created through fully blind mosaicing algorithms, the operation has been performed in black box mode, with no detail provided. Two test scripts have been created to assess the behaviour of the algorithm. The first script (Unit test) assesses the MSF on a known imageset and with a fixed Reachability, Sensitivity value pair. The values must be manually defined upon image contents. The second script (Global test) assesses the MSF on a known imageset looking for the best couple of Reachability and Sensitivity values that maximizes the MSF result, producing a convex solution. At each execution of the algorithm, a new MSF has been computed, with increased Reachability and decreased Sensitivity at a fixed step. The structural distortions have been simulated through application of Barrel filters on fixed sized patches to reproduce a structural loss in parallelism of the image contents.

The execution of the first test on a fixed sized distortion obtained through a Barrel geometrical transform on a 32x32 pixels region and the 10% highest ranked response points on the selected sample image gives an unbiased MSF of 0.0. The execution of the same test on the 10% highest ranked Reachability on the same image sample gives an unbiased MSF of 0.592. This behaviour suggests the capacity of detecting distortions in a region compatible to the maximum size of the SURF points. This solution will not be optimal. The first test is not reliable, as the result is driven by the parameters chosen by the operator.

The second test, executed on the entire set of possible Reachability, Sensitivity values, will find an optimal MSF value of 3.770. A more extensive test has been conducted to assess the response of the algorithm to a set of known photometric and geometrical distortions. An imageset composed of 4 aerial hi-resolution and ISPRS satellite images has been chosen. A set of 64x64 pixel patches has been chosen in a shared textured region of the narrow-view image 2. The MSF measures the



Figure 6. 10% highest response SURF ranked points

extent of actual distorted contents transferred to the mosaic. Luminance distortion (steps -50, -20, -10, -5, -2, 0, 2, 5, 10, 20, 50) has been applied along with Gaussian Blur (normalized steps 0.1, 0.2, 0.3, 0.4, 0.5). The geometrical coherence has been assessed through application of a Barrel distortion (steps 5%, 10%, 20%, 50%). The mean and variance have been computed for each distortion. Table 1 shows the analytic results of the test for each type of distortion and patch size.

The test shows how the SURF points are not invariant to changes in luminance (Figs. 10 and 13) and how tiny changes can have an impact on the position of the SURF points that appear to be more affected by luminance increases. The Gaussian Blur produces a sharp degradation for low values and stabilizing at 0.3, a level where the loss of detail cannot be recovered anymore by the SURF pyramidal filter. The mean (Figure 11) rises and the variance drops beyond 0.2 and stabilizes after 0.3 (Figure 14). The barrel filter (Figure 12 and 15) steadily increases showing capacity of the algorithm to monotonically follow moderate distortion growth. Nevertheless the algorithm shows its limits with Barrel distortions beyond 20%, where sudden changes in mean and variance are experienced. This value corresponds to a distortion whose extent is measured as a different geometry in the scene, yielding to a sudden sharp rise in variance. The results of the tests suggest how the algorithm has a reliable response to geometrical distortions, up to a given extent, but how changes in illumination may severely rise the level of uncertainty as there would not be a way to distinguish geometrical and photometrical contribution to the global MSF score. Gaussian blur loss suggests how image unsharpness and overall loss of detail, may not allow SURF to properly detect reliable points, making MSF unsuitable.

5. CONCLUSION

The algorithm presented in this paper is a first attempt to define a universal metric for Mosaicing Structural Fidelity Assessment (MSF). The metric has been designed to measure the extent of structural inconsistencies caused by geometrical distortions in image sequences and transferred to a mosaic.



Figure 7. 10% highest response reachability SURF ranked points

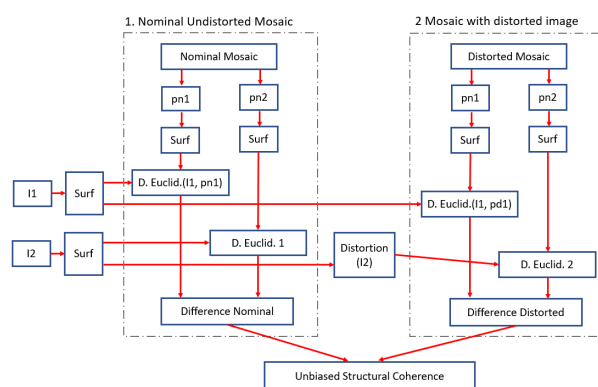


Figure 8. Pipeline of the MSF algorithm

obtained by combining the images of the sequence. The metric can be coupled to an existing mosaicing pipeline to measure the effective geometrical coherence of its internal geometry in applications where GCP georeferencing is not possible or where objects moving within the scene would lead to poor geometrical coherence and cause double contours in mosaics.

The algorithm proposed in this paper estimates the MSF distance through SURF point based Euclidean distance measurements. The structural coherence is analyzed only in the regions where SURF points can be extracted. Uniform regions are considered irrelevant for any MSF estimation. A double approach sensitivity to Reachability is used to detect distortions that may be far from textured regions, or select only distortions in highly textured regions featuring photometrical and geometrical boundaries. The tests show how the algorithm response is driven by photometrical changes and geometrical distortions. The algorithm relies on the invariance of the SURF points to saturation and noises such as the Gaussian Blur.

A test conducted through application of a Barrel distortion on different sized patches shows how small geometrical distortions are properly detected, until a certain extent. Nevertheless the response is not reliable when the image is not sharp and changes

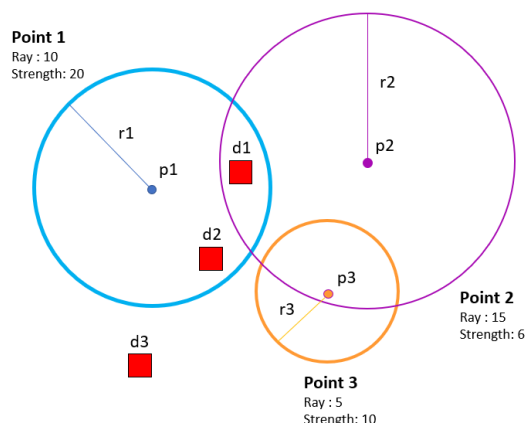


Figure 9. SURF geometrical distortion detection

Table 1. Table of the analytic results of the test for each type of distortion

Patch size	Distortion	Strength	Mean MSF	Var MSF
64x64	Luma	-50	15.041	36.523
64x64	Luma	-20	17.275	58.922
64x64	Luma	-10	13.712	51.520
64x64	Luma	-5	13.772	47.925
64x64	Luma	-2	14.714	98.031
64x64	Luma	0	0.000	0.000
64x64	Luma	2	14.645	158.731
64x64	Luma	5	13.248	139.593
64x64	Luma	10	16.335	85.214
64x64	Luma	20	21.104	132.708
64x64	Luma	50	17.805	150.734
64x64	Gaus. Blur	0.1	8.036	61.303
64x64	Gaus. Blur	0.2	10.312	81.681
64x64	Gaus. Blur	0.3	13.024	24.006
64x64	Gaus. Blur	0.4	13.287	24.724
64x64	Gaus. Blur	0.5	13.286	23.057
64x64	Barrel	1%	5.922	20.309
64x64	Barrel	2%	6.636	21.339
64x64	Barrel	5%	11.294	42.633
64x64	Barrel	10%	11.699	33.739
64x64	Barrel	20%	19.655	200.214
64x64	Barrel	20%	19.454	194.716

in illumination occur. The largest distortions will penalize highly textured regions, where the respect of the geometrical consistency is more important. Future work will aim at providing better immunity to photometrical changes. The capability to distinguish pure geometrical distortions in the same scene to changes in image contents (new objects causing occlusion), may be a further improvement of the solution.

REFERENCES

- Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110, 346–359.
- Bevilacqua, A., Azzari, P., 2006. High-quality real time motion detection using ptz cameras. *IEEE International Conference on Video and Signal Based Surveillance*.
- Boutellier, J., Silvén, O., Tico, M., Korhonen, L., 2008. Objective evaluation of image mosaics. *International Conference on Computer Vision and Computer Graphics*, 107–117.
- Ghosh, D., Kaabouch, N., 2016. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34, 1–11.
- Ghosh, D., Park, S., Kaabouch, N., Semke, W., 2012. Quantitative evaluation of image mosaicing in multiple scene categories. *IEEE International Conference on Electro/Information Technology*.
- 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.
- Li, L., Yao, J., Lu, X., Tu, J., Shan, J., 2016. Optimal seamline detection for multiple image mosaicking via graph cuts. *ISPRS Journal of Photogrammetry and Remote Sensing*, 113, 1–16.
- Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 91–110.
- Rane, K.R., Limbachiya, M.B., Pandit, S.S., Bhirud, S.G., 2003. Mosaic evaluation: An efficient and robust method based on maximum information retrieval. *International Journal of Computer Applications*, 73, 22–26.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13, 600–612.
- 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.

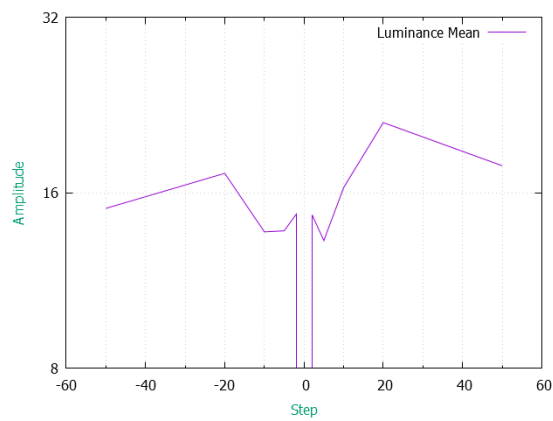


Figure 10. Plot luminance mean 64x64 px

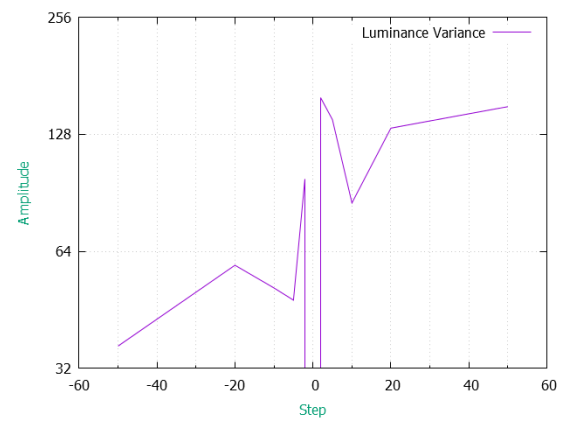


Figure 13. Plot luminance variance 64x64 px

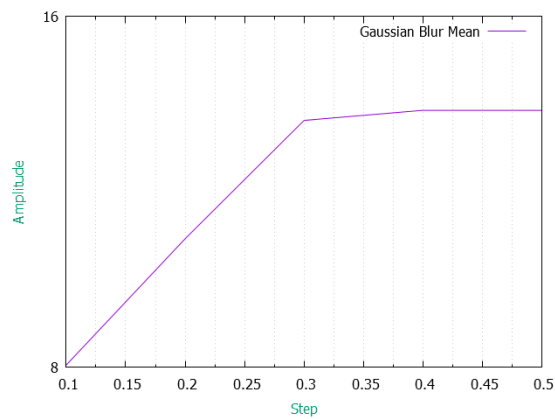


Figure 11. Plot Gaussian Blur mean 64x64 px

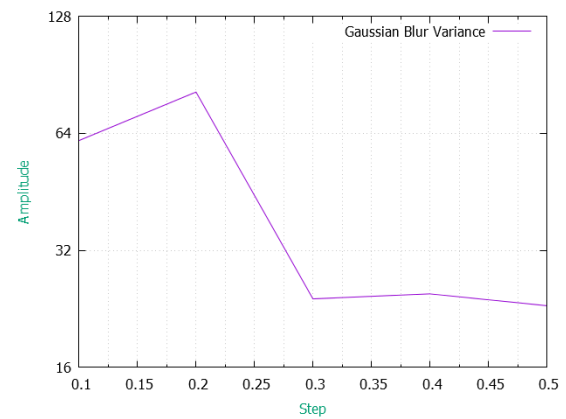


Figure 14. Plot Gaussian Blur distortion variance 64x64 px

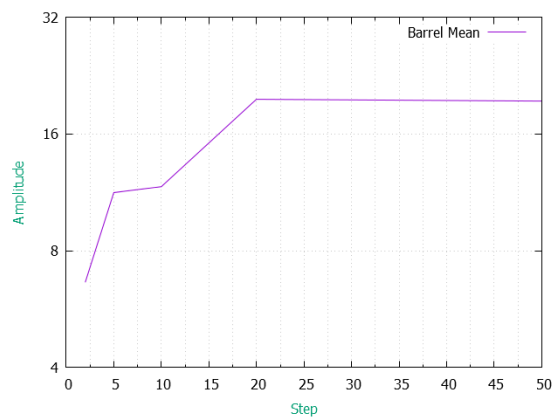


Figure 12. Plot barrel distortion mean 64x64 px

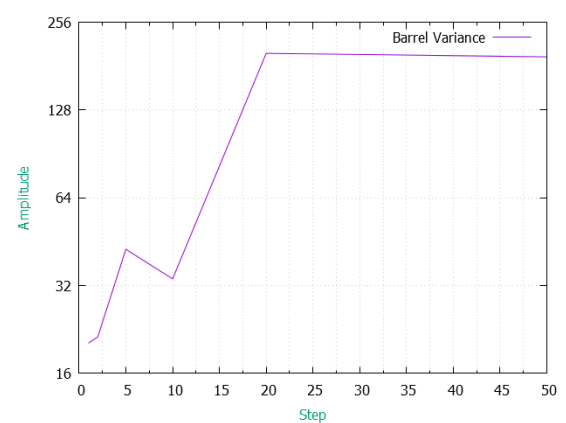


Figure 15. Plot barrel distortion variance 64x64 px