An Advanced Benchmarking for Image Compositing Evaluation
Roberto Giudici, Luc Courtrai, Sébastien Lefèvre

To cite this version:

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 misalignements 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 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 (GIS) 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 (Worrall 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. (Dijkgra, 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, 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, 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 (SSIIM) 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.

(Revilacqua, 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 lighting 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 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 \( T = \{t_0, \ldots, t_{n-1}\} \), where \( nt \) is the number of test cases. For each test case, we assume a set \( S \) of 24-bit RGB input images of size \( n m \) pixels, with \( S = \{x_0, \ldots, x_{nm-1}\} \) and \( n s \) the size of the image set. A pair of consecutive images with \( x_{k-1} \) and \( x_k \) the first and second images respectively, with \( k \in [0, n-1] \). \( \Omega = \{\omega_0, \ldots, \omega_{nm-1}\} \) is the set of mosaic images and the set of mosaic operators is written \( M = \{m_0, \ldots, m_{nm-1}\} \), 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 \( D = \{d_0, \ldots, d_{nd-1}\} \), with \( x_\prime_k = d_j(x_k) \) 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 \( \omega_{ref} = m_{ref}(x_{k-1}, x_\prime_k) \) and the probe mosaic with the algorithm under probe using \( \omega_{probe} = m_{i}(x_{k-1}, x_\prime_k) \). 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 \( S_w = x_{k-1} \cap x_\prime_k \). 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 \( S_{x_{k-1}} \) 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 \( x_\prime_k \) image in the resulting mosaic, leading to image losses. We decided to add the \( S_{x_{k-1}} \) case computed with the fully distorted image \( x_\prime_k \) and average the results to improve reliability. Our approach computes a PSNR and SSIM measuring the difference of contents in shared regions generated by \( S_{x_{k-1}} \) and \( S_{\omega_{probe}} \):

\[
PSNR_{ref} = \frac{1}{2} \left( \frac{PSNR(S_{\omega_{ref}}, S_{\omega_{probe}})}{PSNR(S_{\omega_{ref1}}, S_{\omega_{probe}})} + PSNR(S_{\omega_{ref2}}, S_{\omega_{probe}}) \right)
\] (1)
We then define the disturbance function to the current 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:

\[
\begin{align*}
PSNR_{i,j} &= \frac{1}{nk} \sum_{k=0}^{nk-1} PSNR_{i,j,k}ref, \\
SSIM_{i,j} &= \frac{1}{nk} \sum_{k=0}^{nk-1} SSIM_{i,j,k}ref.
\end{align*}
\]

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.
5. Reference ISPRS imgeset: 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 then 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 worst computational performances (18285.4 sec.).

We have introduce the fidelity analysis metric to measure the dependency between image distortion and mosaicing, as well as 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 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


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.


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

Table 1. Test cases

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Geomatics</th>
<th>Photography and media</th>
<th>Structural preservation</th>
<th>Real-Time</th>
<th>Computational time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feathering</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Fastest</td>
</tr>
<tr>
<td>Multiband</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Fast</td>
</tr>
<tr>
<td>Watershed</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Slow</td>
</tr>
<tr>
<td>Graph-Cut</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Slowest</td>
</tr>
<tr>
<td>Voronoi</td>
<td>Yes</td>
<td>Yes</td>
<td>Not good</td>
<td>Yes</td>
<td>Fast</td>
</tr>
</tbody>
</table>

Table 2. Table of mosaicing use cases


Rosten, E., 2018. FAST corner detection.


