Big Snapshot Stitching with Scarce Overlap

Alexandros-Stavros Iliopoulos* Jun Hu* Nikos Pitsianis†* Xiaobai Sun*
Michael Gehm‡ David Brady§

*Department of Computer Science
Duke University
Durham, NC 27708, USA

†Department of Electrical & Computer Engineering
Aristotle University of Thessaloniki
Thessaloniki 54124, Greece

‡Department of Electrical & Computer Engineering
University of Arizona
Tucson, AZ 85721, USA

§Department of Electrical & Computer Engineering
Duke University
Durham, NC 27708, USA

***Abstract*—We address certain properties that arise in gigapixel-scale image stitching for snapshot images captured with a novel micro-camera array system, AWARE-2. This system features a greatly extended field of view and high optical resolution, offering unique sensing capabilities for a host of important applications. However, three simultaneously arising conditions pose a challenge to existing approaches to image stitching, with regard to the quality of the output image as well as the automation and efficiency of the image composition process. Put simply, they may be described as the sparse, geometrically irregular, and noisy (S.I.N.) overlap amongst the fields of view of the constituent micro-cameras. We introduce a computational pipeline for image stitching under these conditions, which is scalable in terms of complexity and efficiency. With it, we also substantially reduce or eliminate ghosting effects due to misalignment factors, without entailing manual intervention. Our present implementation of the pipeline leverages the combined use of multicore and GPU architectures. We present experimental results with the pipeline on real image data acquired with AWARE-2.**

I. Introduction

A. Image stitching

Digital image stitching offers a means of forming images over an extended field of view (FoV) without sacrificing image resolution. Constituent images may be acquired by a special-purpose camera such as those of planetary rovers [1], orbiting satellites [2], a sky-gazing telescope, a single camera on a robotic mount [3], [4], or a commodity camera with sweep-mode functionality.

Despite great advances in available stitching software, obtaining high-quality mosaics largely relies on several favorable conditions: the captured scene being almost stationary; overlap between adjacent images being large and not adversely corrupted by noise; and calibrated information on the extrinsic and intrinsic parameters of the image acquisition system [5] being fully or partially accessible. Should no information be available on the extrinsic parameters, one may still recover the spatial geometry for a stationary mosaic, provided that the overlaps are rich in distinctive features [6]. Alternatively, the use of camera arrays [7]–[11] may relax or remove the stationary-scene requirement, as well as constrain the space of extrinsic parameters.

AWARE-2, a novel micro-camera array developed by Brady *et al.* [11], allows capturing *snapshots* at gigapixel scale. The distributed FoVs of simultaneously firing micro-cameras can be combined by means of image stitching to form a single high-resolution image over an extended FoV, while foregoing scene motion or other dynamic effects. Successors to AWARE-2 are expected to provide a cost-effective alternative to very high-end or bulky gigapixel cameras. Such unique imaging capabilities have many potential applications, which include wildlife habitat monitoring [12], celestial exploration [1], and recognition or tracking of moving objects or people in crowded scenes [13]. In the rest of this paper, we focus on image stitching with regard to AWARE-2 characteristics, some of which are shared by its precursors, contemporary peers, as well as successors in the making.

B. The S.I.N. conditions

From a computational viewpoint, the constituent AWARE-2 snapshots can be treated as if shot in succession by a single camera capturing an *effectively* static scene. An additional advantage is the absence of parallax between adjacent shots, owing to the optical design of AWARE-2. Nonetheless, design considerations for maximizing the composite FoV and resolution give rise to three challenging conditions for the stitching process.

These conditions pertain to the sparse, geometrically irregular, and noisy (S.I.N.) overlap amongst the constituent FoVs—this is illustrated in fig. I.1: (i) Overlap among constituent images is scarce; any two adjacent images overlap over only a small portion of their combined domain, if at all. (ii) The extrinsic parameters of each micro-camera are defined by two angular rotations [14] and sensor displacement. These geometric parameters are highly irregular across the set of micro-cameras, as a result of manufacturing deviations from the composite design, as well as camera packing on the dome-like mount. (iii) Image data in regions of overlap are highly noisy, due to adverse vignetting and stray

978-1-4799-1365-7/13/$31.00 ©2013 IEEE

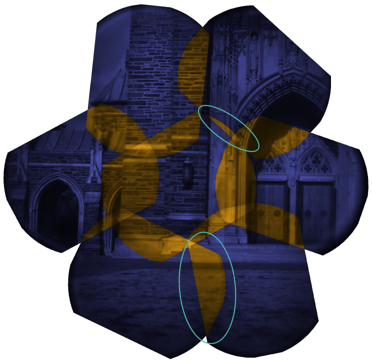


Fig. I.1: A mosaic of 7 AWARE-2 micro-camera images. Overlapping regions are highlighted in orange, and challenging cases are circled. *Top:* overlap is too small. *Bottom:* the region is very poor in distinctive features.

light effects [14]; this becomes more problematic for regions that are lacking in features.

Existing stitching software solutions are considerably challenged by the S.I.N. conditions. Some require the FoV configuration to follow the conventional pattern of a regular grid. On the other hand, grid-free solutions rely on significant and distinctive overlap between neighboring images, yielding poor results with sparsely overlapped ones. These problems are further aggravated by the presence of noise and photometric aberrations.

C. *The stitching pipeline*

Our aim is to obtain appropriate image transformations to effect correct and coherent alignment of the constituent images. Misalignment gives rise to visible geometric and photometric inconsistencies, and often produces ghosting effects—see figs. IV.1a and IV.1c. It stems from deviations in camera location and orientation from the ideal, designed geometry, but also from the inhomogeneous and dynamic nature of camera settings. Moreover, we are concerned with the computational efficiency and scalability of the stitching process. Based on the above considerations, we introduce a computational pipeline, which does not require any manual intervention—see the diagram of fig. I.2. With it, we have obtained satisfactory results in snapshot stitching with AWARE-2 data.

The underlying idea is as follows: We exploit scene-dependent information (features) together with system-specific information. The latter kind, particularly the distributed camera placement configuration, need not be static, but may be subject to subtle changes over time for reasons such as mechanical or thermal drift [4], [7], [14]. In our stitching approach, we use information from extracted features to refine the system information; and we use the designed or calibrated (by experimental or computational means) geometric configuration to confine and guide the feature-based alignment process, or to provide a fallback option should scene-dependent information be insufficient.

We introduce two key stages in sections II and III, demonstrate experimental results in section IV, and conclude the paper with section V.

II. Multiple Image Alignment

The multiple image alignment procedure consists of two basic phases: (i) adjacent images are registered in a pairwise manner, and (ii) the pairwise registration transformations are adjusted into a coherent set of simultaneous transformations. We refer to the latter phase as *bundle adjustment*, notwithstanding the term generally implies 3D reconstruction, which we circumvent as we are interested in producing 2D snapshot mosaics.

A. *Pairwise image registration*

Two images $\mathbf{I}_i$ and $\mathbf{I}_j$ are termed adjacent if they overlap or share a common boundary. Pairwise registration refers, then, to the process of determining a geometric transformation such that, in the absence of noise and photometric variation,

$$\mathbf{I}_i(\mathbf{v}) = \mathbf{I}_j(\mathbf{H}_{ij}\mathbf{v}) \qquad \forall \mathbf{v} \in \mathcal{D}_i \cap \mathcal{D}_j, \tag{II.1}$$

where $\mathbf{v} = [x \quad y \quad 1]^\top$ is a vector of homogeneous coordinates; $\mathcal{D}_i$ indicates the domain of image $\mathbf{I}_i$; and $\mathbf{H}_{ij}$ is the 3×3 transformation matrix.

Automatic and robust image registration methods often resort to feature-based approaches. These entail extracting sets of corresponding features in both images and using them as control points, $\mathbf{v}$, for determining the transformation matrix, $\mathbf{H}_{ij}$. This process is complicated by the presence of noise and photometric variation [4], [15], which it must be made robust for. There is a rich body of literature on feature-based image registration—one may refer to [5] for an elaborated overview—and many different types of features to consider. We make use of SIFT keypoints [16], which are robust for a broad class of transformations, albeit computationally more expensive than others [17]; we use the SiftGPU library [18] to compute the keypoints efficiently.

Because of the S.I.N. conditions, we extend the nominal regions of pairwise overlap, $\mathcal{D}_i \cap \mathcal{D}_j$, to complement imprecise knowledge of the system geometry, and to avoid unnecessary truncation of the domain of potential features. This extension depends on deviation tolerance settings and is refined through data learning.

Feature matching provides control points to be used in determining the registration transformations. It is realized in two steps: descriptor matching and outlier removal. The former is effected using the approach suggested by Lowe in [16], aided and accelerated by system-specific information. The latter is discussed in the next subsection, and it is central to our pipeline.

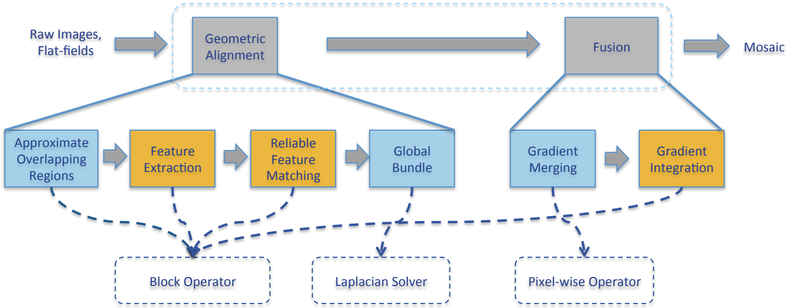


Fig. I.2: A simplified diagram of the image stitching process. Computation-intensive modules are highlighted in orange. High-performance operation categories are indicated by the dashed arrows.

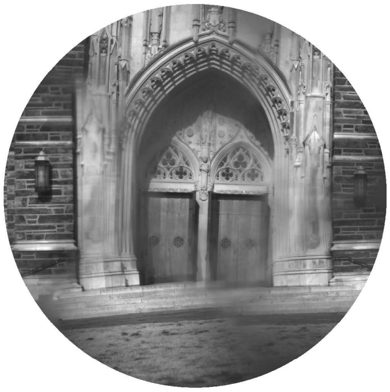


Fig. II.1: Minimal FoV overlap may lead to the estimation false-positive registration transformations.

B. PG-RANSAC

The set of matched descriptors may still contain several false matches (or "outliers"). Such erroneous control points can lead to inconsistent image alignment, or *broken* mosaics, as shown in fig. II.1. Hence, outliers must be removed. This is typically carried out by the RANSAC algorithm [19] or a variation thereof. We have developed a RANSAC-like algorithm, which we call PG-RANSAC, for robust stitching under the S.I.N. conditions. It incorporates the *placement geometry* of multiple cameras, as per their extrinsic parameter ranges, into the outlier removal process. Thus, PG-RANSAC is an integral part of the registration estimation and adjustment process, greatly improving its overall robustness by utilizing system information in tandem with image features.

In order to avoid cases such as that of fig. II.1, we must consider both points and transformations in the outlier removal process, instead of just points. To this end, we augment the ranking function in the RANSAC verification step to make use of the expected placement geometry:

$$\rho'\big(d, \mathcal{T}(\mu), \tilde{\mu}\big) = f\big(\mathcal{T}(\mu), \mathcal{T}(\tilde{\mu})\big) \cdot \rho\big(d, \mathcal{T}(\mu)\big), \quad \text{(II.2)}$$

where d is the feature re-projection distance; $\mathcal{T}(\mu)$ is a transformation $\mathcal{T}$ with parameters μ; $\tilde{\mu}$ refers to the expected parameters; and $\rho(\cdot,\cdot)$ is any RANSAC-style ranking function, normalized to the range $[0, 1]$. The weight function $f(\cdot,\cdot)$ penalizes transformations that deviate much from the expected one, and has the form

$$f(\mathbf{x}, \tilde{\mathbf{x}}) = \prod_{i=1}^{N} \frac{1}{1 + e^{-\alpha[(x_i - \tilde{x}_i) - \tau_i]}} \cdot \frac{1}{1 + e^{\alpha[(x_i - \tilde{x}_i) - \tau_i]}}, \quad \text{(II.3)}$$

where N is the number of deviation measurements, such as camera orientation, that are used to validate estimated transformations; $\tilde{x}_i$ is the expected value for the i-th parameter or measurement; and τ_i is the error tolerance that corresponds to $\tilde{x}_i$, which may be related to the camera array manufacturing/mounting process, noise in sensor readings, etc. The weight function for a single measurement is shown graphically in fig. II.2. It features a plateau in the range $[(\tilde{x} - \tau), (\tilde{x} + \tau)]$, and drops sharply at the boundaries, at a rate dependent on α.

Incorporating this geometric constraint entails the comparison of estimated transformations against the expected ones. To do this, we generate a set of points towards the periphery of the image domain and separately apply the expected and estimated transformations to them. Then, the mean angular and magnitude errors of the motion vectors between the respectively transformed points can be computed [20]; these are used as a measure of deviation of the estimated transformation from the assumed configuration. The set of inliers with respect to this augmented measure is input to the subsequent bundle adjustment. Should no transformation be estimated by the PG-RANSAC process for a pair of images, which might be the case given the small and noisy nature of overlap between images, then a set of 4 anchor points is generated. These effectively guide the bundle adjustment process to respect the expected transformation for the pair.

C. Bundle adjustment

Bundle adjustment generally pertains to the structure-from-motion problem, which implies reconstruction of the 3D scene [21]. Our objective is, rather, to determine a set of projective transformations, one for

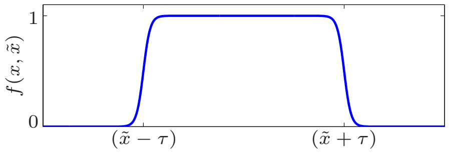


Fig. II.2: Graph of eq. (II.3) for a single deviation parameter.

each constituent image. Still, formulation of this process is analogous to that of a bundle adjustment problem.

Assume that each constituent sub-scene can be captured approximately by a plane, which is reasonable when the distance between the captured scene and the lens is large compared to the focal length of the lens. We may then model the projective relations as homography transformations, and use one of the images as a common reference plane. We form a least-squares (LS) error model for the simultaneous projections:

$$\min_{\{\mathbf{H}_i\}} \sum_{\mathcal{D}_i \cap \mathcal{D}_j \neq \emptyset} \sum_{\mathbf{v}_k \in \mathcal{M}_{ij}} w_{ij} \left\| \mathbf{H}_i \mathbf{v}_{ki} - \mathbf{H}_j \mathbf{v}_{kj} \right\|_2, \quad \text{(II.4)}$$

where $\mathcal{M}_{ij}$ denotes the set of matched control points between two adjacent images; $\mathbf{v}_{ki}$ is the local homogeneous coordinate vector of the k-th control point in $\mathcal{D}_i$; and w_{ij} is a weight that normalizes the contribution of each pair to the solution.

We have developed a fast method for minimizing the above error. The LS solution to eq. (II.4) can be obtained through the corresponding system of normal equations. The related system matrix is a block-wise sparse, with a 3×3 block for each image pair. Note that its size depends solely on the number of constituent images, and not on the number of control points. The block structure corresponds directly to the Laplacian matrix of an adjacency graph where nodes and edges correspond to cameras and regions of FoV overlap, respectively—see fig. II.3. The block-Laplacian matrix can be constructed easily and in parallel. Next, by choosing a reference plane among the participating image planes, we recast the homogeneous normal equations to a non-homogeneous system, circumventing the null space problem and short-cutting an otherwise lengthy and complex iteration process [21], [22]. The solution to the final, linear system can be obtained directly and efficiently. The fast process is robust because of the PG-RANSAC framework providing reliable control points.

III. Image Fusion

Following the bundle adjustment process, every pixel can be projected onto a specific location of the mosaic canvas, albeit different pixels might share the same location. Photometric variation between overlapping and adjacent image regions may result in visible seams in the stitched image [23]. In order to fuse the overlapping image data, we blend the images in the gradient domain [24], [25]. This technique has the advantage of smoothing intensity seams, while maintaining high-frequency information in the source images; moreover, image gradients are invariant to camera sensor bias.

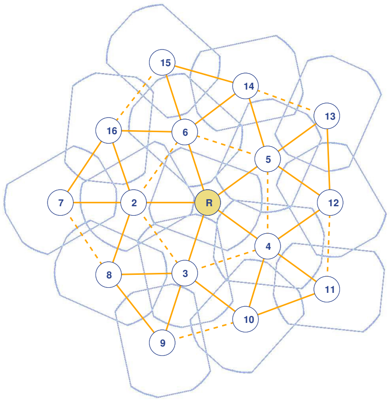


Fig. II.3: A portion of the AWARE-2 camera adjacency graph. Edges in dashed lines indicate very small overlap. The plane associated with node R is chosen as the reference plane.

Let $\hat{\mathbf{I}}_c$ be the c-th projected image on the mosaic canvas. Gradient-domain blending dictates that the blended mosaic, $\hat{\mathbf{I}}$, must satisfy

$$\nabla \hat{\mathbf{I}}(\mathbf{v}) = \sum_{\hat{\mathcal{D}}_c \ni \mathbf{v}} w_c(\mathbf{v}) \nabla \hat{\mathbf{I}}_{\mathbf{c}}(\mathbf{v}), \quad \text{(III.1)}$$

where $w_c(\mathbf{v})$ is a weighting function that we set to be the reciprocal of the pre-calibrated flat-field measurement for each pixel $\mathbf{v} \in \hat{\mathcal{D}}_c$, which reflects its apparent gain and dark current [14]. The intensity-domain mosaic is then computed via numerical integration.

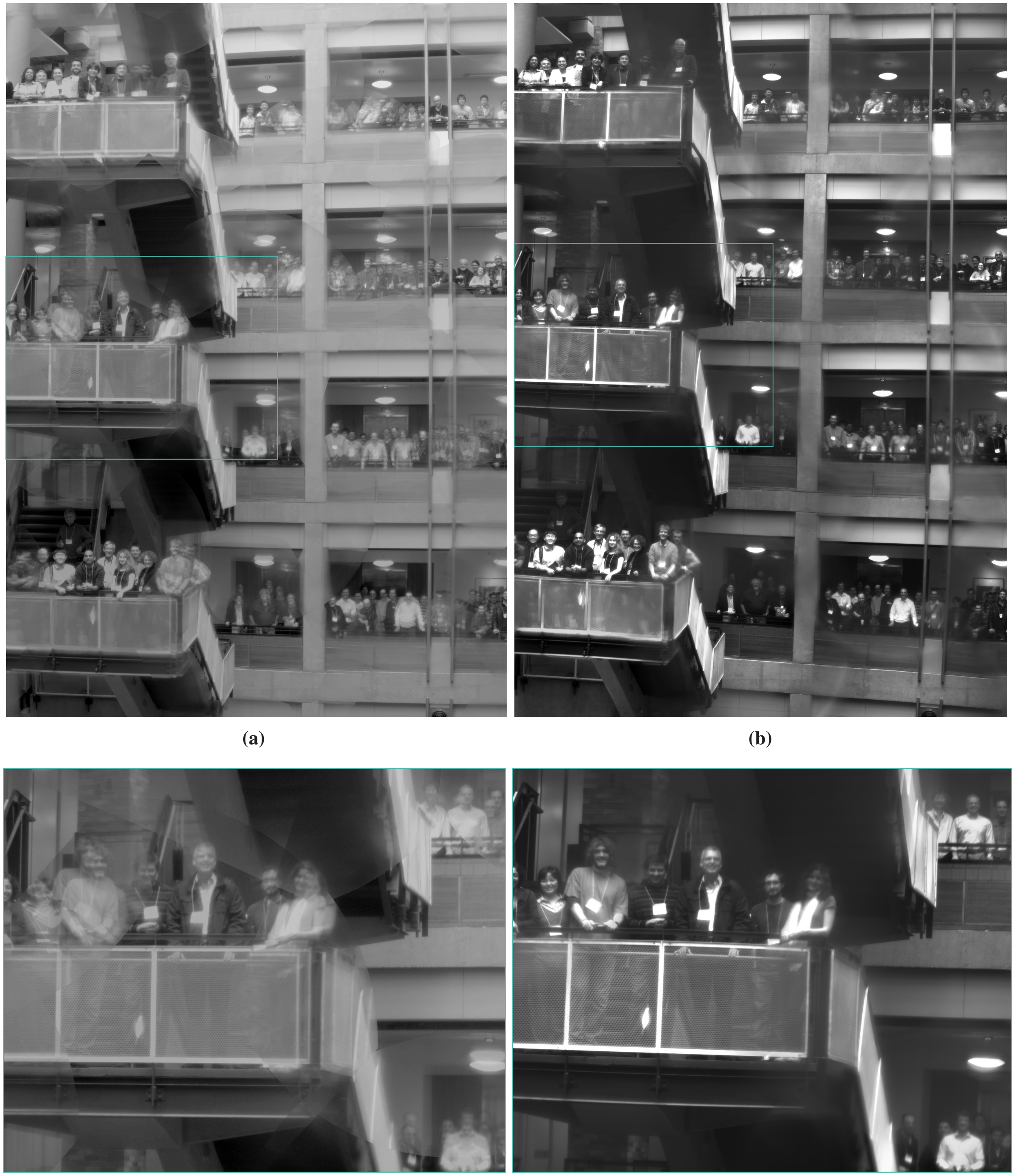
We perform the image fusion computations entirely in the GPU. Constituent images are back-projected using bi-linear interpolation, and are subsequently differentiated in batches of non-overlapping images. Last, we employ the convolution pyramid scheme [26] to effect a fast approximation of the integration operation.

IV. Demonstration

Our pipeline implementation makes combined use of multicore and GPU architectures. Representative experimental results, obtained with our pipeline on real data acquired with AWARE-2, are shown in fig. IV.1. The images to the right were produced automatically and processed within half a minute, without overlapping CPU-GPU data transfers and computational processing.

V. Additional Remarks

We have outlined a computational pipeline for image stitching at gigapixel scale, with scarce overlap of the constituent fields of view, and under the presence of noise, photometric variation, as well as other factors that make the image acquisition conditions deviate from the ideal, designed configuration. While existing



(a)

(b)

(c)

(d)

Fig. IV.1: Snapshot mosaics of a live scene, captured with the AWARE-2 camera prototype at the International Conference on Computational Photography, Seattle, 2012 (bottom floor and architectural surroundings not shown). *(a),(c):* Results produced by the AWARE-2 compositing pipeline [14], where tone mapping has been applied. *(b),(d):* Results produced automatically by the alternative pipeline introduced in this paper, without tone mapping. *Top row*: The displayed scene spans the fields of view of approximately 25 micro-cameras. *Bottom row:* Detail, zoomed in within the marked windows in the top-row images.

software for image stitching fails or performs poorly with AWARE-2 data, our pipeline in its present implementation succeeded in effecting robust and efficient stitching of high-resolution, ghost-free mosaics. In order to achieve this, we explore and leverage the acquisition conditions in tandem with image features, while using one to refine the other. Further, we exploit advanced algorithm techniques and modern parallel architectures to automate and accelerate the stitching process. While use of a snapshot camera array eliminates scene motion effects, efficiency in the stitching process will become paramount for applications using such camera systems for image or video analysis of dynamically evolving scenes, by enabling the monitoring of many kinds of scientific or social phenomena at fast-changing rates.

The pipeline can be easily ported to other applications, not restricted to snapshots or AWARE-2. This paper is the first written document about its overall structure.[1] Due to space constraints, detailed descriptions and potential improvements are omitted and will be reported elsewhere.

The authors wish to thank Esteban Vera Rojas, Steve Feller, Daniel Marks, Lars Nyland, Changchang Wu, as well as our anonymouns reviewers.

References

[1] M. R. Balme, A. Pathare, S. M. Metzger, M. C. Towner, S. R. Lewis, A. Spiga, L. K. Fenton, N. O. Renno, H. M. Elliott, F. A. Saca, T. Michaels, P. Russell, and J. Verdasca, "Field measurements of horizontal forward motion velocities of terrestrial dust devils: Towards a proxy for ambient winds on mars and earth," *Icarus*, vol. 221, no. 2, pp. 632–645, 2012.

[2] D. Antony and S. Surendan, "Satellite image registration and image stitching," *International Journal of Computer Science & Engineering Technology*, vol. 4, no. 2, pp. 62–66, 2013.

[3] M. Ben-Ezra, "A digital gigapixel large-format tile-scan camera," *IEEE Computer Graphics and Applications*, vol. 31, no. 1, pp. 49–61, 2011.

[4] J. Kopf, M. Uyttendaele, O. Deussen, and M. F. Cohen, "Capturing and viewing gigapixel images," *ACM Transaction on Graphics*, vol. 26, no. 3, 2007.

[5] R. Szeliski, *Computer vision: Algorithms and applications*. London; New York: Springer, 2010.

[6] M. Brown and D. G. Lowe, "Recognising panoramas," in *Proceedings of the 9th IEEE International Conference on Computer Vision*, ser. ICCV '03, vol. 2, 2003, pp. 1218–1225.

[7] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics*, vol. 24, no. 3, p. 765, 2005.

[8] R. Horisaki, Y. Nakao, T. Toyoda, K. Kagawa, Y. Masaki, and J. Tanida, "A thin and compact compound-eye imaging system incorporated with an image restoration considering color shift, brightness variation, and defocus," *Optical Review*, vol. 16, no. 3, pp. 241–246, 2009.

[9] B. Leininger, J. Edwards, J. Antoniades, D. Chester, D. Haas, E. Liu, M. Stevens, C. Gershfield, M. Braun, J. D. Targove, S. Wein, P. Brewer, D. G. Madden, and K. H. Shafique, "Autonomous real-time ground ubiquitous surveillance-imaging system (ARGUS-IS)," *Proceedings of SPIE*, vol. 6981, pp. 69 810H–1–69 810H–11, 2008.

[10] D. L. Marks and D. J. Brady, "Close-up imaging using microcamera arrays for focal plane synthesis," *Optical Engineering*, vol. 50, no. 3, pp. 033 205–1–033 205–40, 2011.

[11] D. J. Brady, M. E. Gehm, R. A. Stack, D. L. Marks, D. S. Kittle, D. R. Golish, E. M. Vera, and S. D. Feller, "Multiscale gigapixel photography," *Nature*, vol. 486, no. 7403, pp. 386–389, 2012.

[12] M. H. Nichols, G. B. Ruyle, and I. R. Nourbakhsh, "Very-high-resolution panoramic photography to improve conventional rangeland monitoring," *Rangeland Ecology & Management*, vol. 62, no. 6, pp. 579–582, 2009.

[13] L. Gueguen, M. Pesaresi, and P. Soille, "An interactive image mining tool handling gigapixel images," in *Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium*, ser. IGARSS '11, 2011, pp. 1581–1584.

[14] D. R. Golish, E. M. Vera, K. J. Kelly, Q. Gong, P. A. Jansen, J. M. Hughes, D. S. Kittle, D. J. Brady, and M. E. Gehm, "Development of a scalable image formation pipeline for multiscale gigapixel photography," *Optics Express*, vol. 20, no. 20, pp. 22 048–22 062, 2012.

[15] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[17] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.

[18] C. Wu, "SiftGPU: A GPU implementation of scale invariant feature transform (SIFT)," 2007, [Online] Available at: http://cs.unc.edu/~ccwu/siftgpu/.

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[20] D. Robinson and P. Milanfar, "Fast local and global projection-based methods for affine motion estimation," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 35–54, 2003.

[21] Y. Jeong, D. Nistér, D. Steedly, R. Szeliski, and I.-S. Kweon, "Pushing the envelope of modern methods for bundle adjustment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 8, pp. 1605–1617, 2012.

[22] M. I. A. Lourakis and A. A. Argyros, "SBA: a software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software*, vol. 36, no. 1, pp. 1–30, 2009.

[23] A. Eden, M. Uyttendaele, and R. Szeliski, "Seamless image stitching of scenes with large motions and exposure differences," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ser. CVPR '06, vol. 2, 2006, pp. 2498–2505.

[24] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," in *Computer Vision – ECCV 2004*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, vol. 3024, pp. 377–389.

[25] A. Agarwala, "Efficient gradient-domain compositing using quadtrees," *ACM Transaction on Graphics*, vol. 26, no. 3, 2007.

[26] Z. Farbman, R. Fattal, and D. Lischinski, "Convolution pyramids," *ACM Transaction on Graphics*, vol. 30, no. 6, pp. 175:1–175:8, 2011.

[1] An initial version was orally presented at the GPU Technology Conference, San Jose, CA, USA, March 2013: http://nvidia.fullviewmedia.com/gtc2013/0320-210A-S3219.html