

GPU-ACCELERATED CROSS-MODAL SAR-EO IMAGE HOMOGRAPHY ESTIMATION

Christopher Beam, Andrew Willis

University of North Carolina at Charlotte
Department of Electrical and Computer Engineering
Charlotte, NC 28223-0001
Email: {cbeam18, arwillis}@uncc.edu

Garrett Demeyer, Kevin Brink

Air Force Research Laboratory
Munitions Directorate
Eglin AFB, FL, 32542
Email: garrett.demeyer.1@us.af.mil,
kevin.brink@us.af.mil

ABSTRACT

This article proposes a massively-parallel approach for solving the difficult problem matching high-altitude image pairs from different domains, e.g., Synthetic Aperture Radar (SAR) and visible light imagery (EO). The through-weather measurement capability of SAR allows this technology to yield vehicle position fixes in inclement weather and during either night or daytime for image-aided navigation. This work focuses on developing capabilities to match across a large range of variations in the unknown parameters of the homography that brings these image pairs into correspondence. This is a problem that is not well-solved by any existing approaches and is important in practice as cross-domain imagery from aerial platforms often exhibits large variations in scale, key-stone, rotation and translation effects that can be different in the x and y axes. Our approach for cross-modal image matching uses a mutual information loss function and applies a massively-parallel search procedure in CUDA to detect and explore the loss function to find satisfactory homographies to match the image pairs. Experiments are performed using simulated image telemetry obtained by flying a fixed wing aircraft in a virtual environment with image data derived from Google Maps and RADARSAT Google Earth Engine image databases. Results show a comparison rate of 12.79 Gpixel/sec and has a search rate of 1.8M matches/sec allowing for exhaustive search solutions. Our approach is found to yield accurate homography values according to our normalized corner error metric for 68% of the image database pairs.

1. INTRODUCTION

Image matching technologies are critical for a large number robotics, computer vision and artificial intelligence applications including odometry estimation, 3D mapping and image based classification. This article considers aerial imagery in the form of image pairs that originate from aerial platforms having different modalities with a focus on Synthetic Aperture Radar (SAR) and visible light imagery (EO).

Our application for this technology is to derive vehicle pose information from SAR-EO image match solutions as alternative to existing position sensors such as an Inertial Navigation System (INS) or a Global Positioning System sensor (GPS). In this context, our algorithm matches a sensed SAR image to a geo-referenced EO image that is already available onboard the aerial platform. Matches between sensed SAR imagery and the reference EO image can provide estimates for vehicle location and pose in GPS-denied contexts. Further, SAR's through-weather measurement capability allows vehicle position fixes to be generated in inclement weather and during either night or daytime since SAR is an active sensing modality unimpeded by rain, clouds, smoke and other atmospheric phenomenon consisting of particles much smaller than the sensing radar wavelength.

While image registration is a classical image processing and computer vision problem, the novelty of this work is: (1) the proposed algorithm finds matches for potentially large variations in the unknown homography parameters, (2) an approach for computing the search parameters to optimize the computational cost is proposed, and (3) a general method for massively-parallel grid search optimization using GPU acceleration is proposed capable of searching over a vast range candidate solutions in a short period of time. For example, our experimentally measured search velocity for an NVIDIA A6000 GPU is 1.8M homographies/sec for a 100x100 image pair.

2. RELATED WORK

Limitations in reception of Global Positioning System (GPS) signals have motivated image aided navigation from remotely sensed images. Specifically, GPS signals can be blocked by obstructions, e.g., tall buildings, which can interrupt the signal or degrade its accuracy [1]. Image-aided navigation methods [1, 2, 3, 4, 5, 6, 7] provide an alternative to GPS. Image-aided navigation systems match sensed imagery to available georegistered imagery using an image registration algorithm and use the correspondence from image registration to esti-

mate absolute position using algorithms such as the Point-and-Perspective (PnP) algorithm. In GPS-denied and GPS-degraded environments these algorithms provide critical absolute position estimates crucial for bounding navigational error by providing noisy GPS-like updates to onboard guidance, navigation and control systems when GPS is unavailable. Approaches to solution include image processing [8], a large-scale image database searches [2] and many other image registration approaches [1, 3, 5].

3. METHODOLOGY

This work considers image matches to conform to an 8-parameter homography which is typical for high-altitude imagery where the height of terrain features is significantly less than that of the altitude of the vehicle. Our approach for solution applies a purpose-built massively-parallel grid search algorithm implemented in CUDA to accelerate the process of finding valid image matching solutions.

The proposed approach consists of three steps: (1) choose an image matching loss function, (2) characterize the loss function's geometric structure, and (3) apply a massively-parallel search procedure to explore the loss function surface and detect minima, i.e., solutions, to the problem by considering for a large/exhaustive set of candidate solutions.

3.1. Choose an image matching loss function

Our cross-modal application that matches EO and SAR images uses a mutual information (MI) image alignment performance functional to measure the quality of a candidate solution as shown in equation (1)

$$I(X; Y | \mathbf{H}) = \sum_{x \in X} \sum_{y \in Y} P_{XY}(x, y | \mathbf{H}) \log_2 \left[\frac{P_{XY}(x, y | \mathbf{H})}{P_X(x | \mathbf{H}) P_Y(y)} \right] \quad (1)$$

where the pair (X, Y) denotes the (SAR, EO) image pair to be matched under the correspondence given by the homography $\mathbf{H}$. MI has been shown to be effective in cross modal alignment of EO-SAR image pairs by other researchers [9, 10] having been initially proposed for cross modal medical image alignment [11]. This objective function is highly non-linear and, if a good initial guess is not provided, methods using approach are not likely to converge to the correct solution.

Our massively-parallel search problem seeks to consider a sparse collection of homographies across the plausible domain of parameter variation to find a solution within the basin of attraction, i.e., geometrically "close", to the correct solution where this alignment metric will converge to the correct solution. While this work uses mutual information, the search procedure will allow any performance functional to be defined and used as a measure of match quality.

3.2. Characterize the loss function's geometric structure

An 8-dimensional parameterization of an image homography is considered. The 8 parameters of the homography are as follows: (1) which include rotation θ , (2,3) non-isomorphic (x, y) scale (α_x, α_y) , (4) shear γ , (5,6) (x, y) translation (t_x, t_y) , and (7,8) (x, y) perspective (keystone) values (k_x, k_y) . The grid search associates each of these parameters to distinct grid axes.

$$\mathbf{H} = \begin{bmatrix} \alpha_x \cos(\theta) & -\alpha_y (\sin(\theta) - \gamma \cos(\theta)) & t_x \\ \alpha_x \cos(\theta) & \alpha_y (\cos(\theta) + \gamma \sin(\theta)) & t_y \\ k_x & k_y & 1 \end{bmatrix} \quad (2)$$

Computational costs for searching the 8-dimensional space is high and GPU acceleration is leveraged to yield solutions at unprecedented rates. Significant computational savings can be afforded by limiting the number of samples along each dimension of the search. For this reason, we characterize the structure of the chosen performance functional (mutual information in this article) to determine appropriate grid sample spacing for each axis of the search.

Our process for characterizing the loss function proceeds by performing image pair matches for different image database pairs with differing degrees of grid resolution. The results of this analysis provides empirical data needed to determine the typical dimensions of the non-linear basin of attraction for the image alignment performance functional in 8-dimensional space. In our case, the basin of attraction will be spanned by each of the 8 dimensions of the homography having the parameters shown in equation 2. We choose a value for the grid sample spacing parameter $\Delta \mathbf{H}$ equal to approximately half the width of the observed shape of the basin of attraction in each dimension.

3.3. Apply a massively-parallel search procedure

A massively parallel GPU-accelerated search is performed over the multi-dimensional grid where each grid point is a possible solution and the quality of each solution is evaluated using the user-defined performance functional (mutual information). Our implementation of MI on GPU follows that in equation (1). One particular challenge for this performance functional is the requirement to compute the joint distribution of intensities for the EO and SAR image. This distribution must be computed on GPU and can require significant memory resources to store when computing +1M image matches. Each GPU thread will be performing the MI calculation on their assigned grid point, leading to memory constraints at the thread level. This requires matches to consider the trade-off between having high resolution joint distributions and a large number of parallel threads.

To help speed up the computation each thread performs some pre-calculation steps to ensure the thread only considers overlapping pixels between the image pair for the given homography. This improves efficiency, especially for large

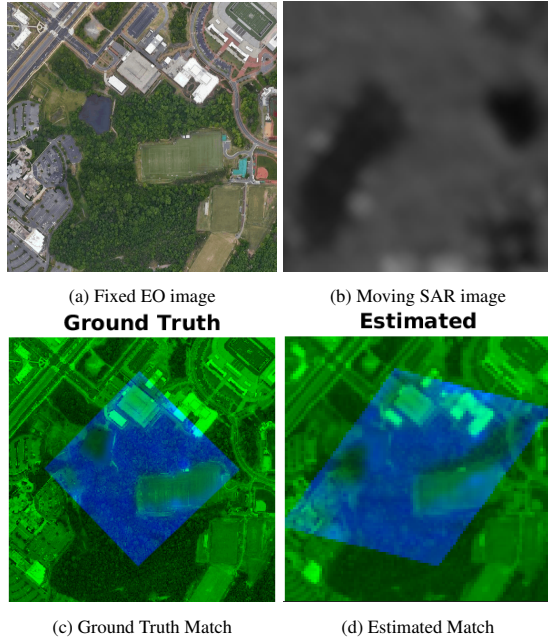


Fig. 1: (a,b) shows an (EO,SAR) image pair from our database, (c,d) show the (ground truth,estimated) image matches with the SAR image overlaid (in blue) on the geo-referenced EO image (in green).

scale values where a single pixel is interpolated and compared with groups of pixels in the matched image. This performance savings is provided by calculating the image of the homography-transformed line corresponding to each pixel raster of the moving image in the fixed image. The parametric equation for this line is intersected with the boundary lines of the fixed image to provide start and stop pixel indices for GPU thread computation. This results in computational savings by avoiding consideration of pixels in the moving image that do not correspond to a pixel inside the boundary of the fixed image. This approach also reveals how much area of each image was used to calculate the MI, allowing for consideration of this property in downstream algorithms.

At the conclusion of the search values of performance functional across the search grid are returned to the host and a minimum value for the performance functional is found using a second GPU min-search which requires little time. This yields a final estimate for the unknown parameters.

4. RESULTS

Our experimental results analyze performance on 538 cross-modal (EO,SAR) image pairs derived from Google Maps and RADARSAT Google Earth Engine image databases respectively where images were recorded from a simulated virtual fixed wing vehicle navigating a spiral-shaped course at an altitude of 500m [12]. Scale variation between Google maps images was approximately have a resolution of 1.2 m./pixel and RADARSAT images with a resolution of approximately

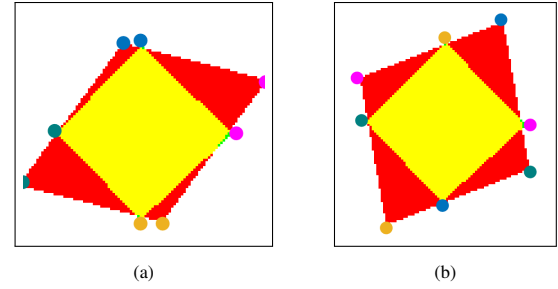


Fig. 2: (a) Acceptable Average Corner Error Example, $\epsilon = 0.1792$, (b) Not Acceptable Average Corner Error Example, $\epsilon = 0.8242$

10 m./pixel of urban Charlotte (Lat,Lon.)= $(-80.741692E, 35.309003N)$ (Urban, 538 pairs). Our search over candidate homographies considered 32 rotations (0° to 348.75°), 256 non-isomorphic (x,y) scale pairs on a 16×16 grid (0.253 to 0.843 for x, 0.256 to 0.861 for y), 5 shear values (-0.4 to 0.2), and 961 candidate translation vectors sampled on a 31×31 grid (-50 to 50). The experiments were performed on a system containing an Intel Xeon Silver 4110, Nvidia RTX A6000 48 GB, and 160 GB RAM which was capable of processing 100x100 image pair matches at a rate of approximately 1.28M image matches per second (12.79 Gigapixels per second). This rate is an approximation as the homographies searched may have 50% to 100% pixel overlap.

A new performance metric is defined to evaluate match accuracy which we refer to as the average normalized corner error (NCE). NCE measures match quality in terms of the average distance of the projected image corners for the ground truth and estimated homographies in images having (x,y) ranges scaled to lie in the $(x,y) \in [0,1]$ as shown in equation (3)

$$\epsilon = \frac{1}{4} \sqrt{\sum_{i=1}^4 (\mathbf{p}_i - \hat{\mathbf{p}}_i)^2} \quad (3)$$

where $(\mathbf{p}_i, \hat{\mathbf{p}}_i)$ denotes the i^{th} point pair for the (4) (x,y) corner locations given by the ground truth and estimated homography respectively.

Figure 1(c,d,e) shows an example (SAR, EO) image match (c,d) and the distribution of the NCE for satisfactory estimates. Figure 2 shows an illustration of the importance of the performance metric, where there is significant overlap (yellow) between the ground truth transformation (green) and the estimated transformation (red), however the estimated corners in (b) are on the opposite side of the ground truth corners.

Using a threshold of $\epsilon < 0.2$ the proposed approach found satisfactory homography estimates for 68% of the tested images pairs. The performance of this algorithm far surpasses other cross-modal approaches explored which included OpenCV feature matching [13] and SIFTFlow [14] which were unable to provide homography estimation results satisfying $\epsilon < 0.2$.

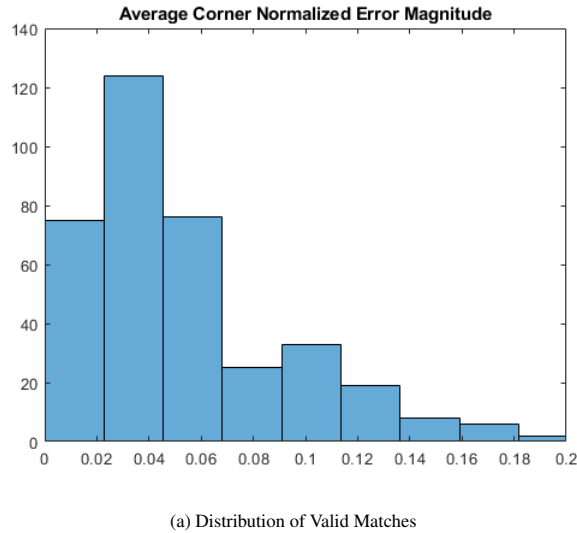


Fig. 3: (Shows the distribution of the normalized corner error (NCE) for our valid matches ($\epsilon < 0.2$).

5. CONCLUSION

An new approach was described to solve the difficult problem of matching image pairs coming from different domains capable of coping with large variations in the unknown parameters of the non-rigid transformation. A purpose-built massively-parallel grid search algorithm is implemented in CUDA to accelerate image matching problems. Mutual information is chosen as the alignment performance functional to measure the candidate solution. Average normalized corner error (NCE) was used as the performance metric to compute the distances of the projected image corners for the ground truth and estimated homographies. Using these methods with a threshold of $\epsilon < 0.2$ for NCE, 68% of the homographies found were satisfactory. The approach presented can be readily extended and improved using adaptive stepsize approaches or other methods for improved search space traversal such as AI-based driven search optimizations.

6. REFERENCES

- [1] Karsten Mueller, Jamal Atman, Nikolai Kronenwett, and Gert F. Trommer, "A multi-sensor navigation system for outdoor and indoor environments," *Proceedings of the 2020 International Technical Meeting of The Institute of Navigation*, 2020.
- [2] D. Venable and J. Raquet, "Large scale image aided navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, pp. 2849–2860, 2016.
- [3] Timothy Machin, John Raquet, David Jacques, and Donald Venable, "Real-time implementation of vision-aided navigation for small fixed-wing unmanned aerial systems," *Proceedings of the 29th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2016)*, 2016.
- [4] Robert C.; Leishman, Jeremy; Gray, John F.; Raquet, and Adam Rutkowski, "Improvements for vision based navigation of small, fixed-wing unmanned aerial vehicles," 2018, Accessed: 2021-8-18.
- [5] Jamal Atman, Manuel Popp, Jan Ruppelt, and Gert F. Trommer, "Navigation aiding by a hybrid laser-camera motion estimator for micro aerial vehicles," *Sensors*, vol. 16, no. 9, 2016.
- [6] Philipp Crocoll, Justus Seibold, Manuel Popp, and Gert F. Trommer, "Indoor navigation for a micro aerial vehicle aided by laser range finder measurements," in *European Navigation Conference, April, 23-25, 2013, Wien, Austria*. 2013, p. 13 S., Austrian Institute of Navigation (OVN).
- [7] Evan T. Dill and Maarten Uijt de Haag, "3d multi-copter navigation and mapping using gps, inertial, and lidar," *NAVIGATION*, vol. 63, no. 2, pp. 205–220, 2016.
- [8] Gianpaolo Conte and Patrick Doherty, "An integrated UAV navigation system based on aerial image matching," *IEEE Aerospace Conference Proceedings*, 2008.
- [9] Aurélien Yol, Bertrand Delabarre, Amaury Dame, Jean-Émile Dartois, and Eric Marchand, "Vision-based absolute localization for unmanned aerial vehicles," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 3429–3434.
- [10] Bhavit Patel, Timothy D. Barfoot, and Angela P. Schoellig, "Visual localization with google earth images for robust global pose estimation of uavs," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 6491–6497.
- [11] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
- [12] Andrew R. Willis, Kevin M. Brink, and Kathleen M. Dipple, "ROS georegistration: Aerial multi-spectral image simulator for the robot operating system," *CoRR*, vol. abs/2201.07863, 2022.
- [13] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [14] Ce Liu, Jenny Yuen, and Antonio Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.