Automatic georeferencing of imagery from high-resolution, low-altitude, low-cost aerial platforms

Amanda Geniviva, Jason Faulring and Carl Salvaggio

Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY, USA

ABSTRACT

Existing nadir-viewing aerial image databases such as that available on Google Earth contain data from a variety of sources at varying spatial resolutions. Low-cost, low-altitude, high-resolution aerial systems such as unmanned aerial vehicles and balloon-borne systems can provide ancillary data sets providing higher resolution, oblique-looking data to enhance the data available to the user. This imagery is difficult to georeference due to the different projective geometry present in these data. Even if this data is accompanied by metadata from global positioning system (GPS) and inertial measurement unit (IMU) sensors, the accuracy obtained from low-cost versions of these sensors is limited. Combining automatic image registration techniques with the information provided by the IMU and onboard GPS, it is possible to improve the positioning accuracy of these oblique data sets on the ground plane using existing orthorectified imagery available from sources such as Google Earth. Using both the affine scale-invariant feature transform (ASIFT) and maximally stable extremal regions (MSER), feature detectors aid in automatically detecting correspondences between the obliquely collected images and the base map. These correspondences are used to georeference the high-resolution, oblique image data collected from these low-cost aerial platforms providing the user with an enhanced visualization experience.

Keywords: image matching, wide-baseline, multi-view imagery, automatic georeference, ASIFT, MSER, oblique

1. INTRODUCTION

Georeferencing of an image is the process of assigning spatial infomation (with respect to an Earth-based coordinate frame) to an image to define its location and orientation. This can be done by registering the image to a geo-accurate base map.[1] Image registration is the ability to transfer two images taken from different times, sensors, or viewpoints to the same coordinate system. Registering the image to a geo-accurate base map will therefore transfer it into the Earth-based coordinate frame. This is commonly accomplished by a user manually selecting control points between the images, a process that can be very time consuming.[1] Automatic image registration requires the feature correspondences to be found without user interaction. There currently exists a number of feature matching algorithms that automatically find and match control points across two images. These algorithms typically locate all distinctive features in each image, use the local neighborhood of the feature to create a feature descriptor, and then find matching control points by matching these descriptors across images.[2] However, these feature detectors tend to have a harder time in the wide-baseline image matching problem. The larger the change in viewpoint, the greater the distortion to the image.[3] This distortion leads to difficulty in finding matching feature descriptors across the two images.

An important step towards fully automatic image georeferencing is the ability to find reliable correspondences between two wide-baseline images taken with different imaging platforms. It's important that these correspondences lie on the ground plane since non-ground plane structures contain strong parallax when imaged from aerial platforms.[4] Finding these correspondences requires an initial estimate of geographic location in order to extract candidate nadir-looking imagery for matching. In order to perform this matching step, a robust feature-correspondence algorithm is needed that is invariant to certain geometric transformations.

This paper proposes a technique for automatic georeferencing of low-altitude, high-resolution oblique imagery. Section 2 briefly discusses previous work on the subject and Section 3 details the specific problem this

Further author information: (Send correspondence to A. Geniviva)
E-mail: amg6210@rit.edu

paper is aimed at solving. Section 4 presents the proposed workflow that will be implemented to perform the georeferencing and additionally details the specifics of the feature matching algorithm being used. Preliminary results are outlined in Section 5, while conclusions and future work are identified in Section 6.

2. PREVIOUS WORK

A fundamental image processing task in the field of remote sensing is the task of image registration. Automatic georeferencing relies on the ability to register an image to one that is georeferenced, thereby transfering the geospatial information. The task of image registration is a difficult task to accomplish automatically due to the occurence of false tie points and the large amount of data that must be processed when dealing with images. While there are still some applications that require an analyst to manually select control points between images, other methods have been developed for specific applications.[2] The wide-baseline problem adds difficulty to this task due to the large amount of image deformation that occurs in oblique imagery.[3]

A very popular solution to this task of image registration is feature detection and matching. Many different algorithms have been developed to accomplish this task. The main categories describing these algorithms are area- and feature-based methods.[5]

Area-based methods rely on the feature matching step alone to find correspondences between images. These methods typically use a window (that could be as large as the image) to find correspondences with another image. Methods such as cross-correlation[6] (CC), Fourier phase correlation,[7] maximization of mutual information[8] (MMI), and optimization methods are used and are reliant on area based feature matching to register images. These methods are useful in certain applications, however, there are limitations that make them unsuitable for this application. The main problem is that they are not effective on images with any type of deformation caused by varying viewpoint. Additionally, using a window to find a corresponding area doesn't guarantee that there will be any remarkable or distinctive feature in the automatically chosen window.[5] Both of these reasons make these methods poorly suited for this application.

Feature-based methods are typically a more versatile approach. A typical feature-based method consists of a feature detector and descriptor. During the feature detection stage, the algorithms locate points of interest in the image. The region around these points is used to generate the feature descriptor. These descriptors are then matched across images in order to locate control points.[2] In evaluations of local feature descriptors, the Scale-Invariant Feature Transform (SIFT) has shown to be one of the best, even in cases with modest changes in viewpoint.[9] However, the SIFT algorithm fails with large changes in viewpoint due to the fact that it does not consider the camera axis orientation paramaters at all. In these wide-baseline problems, a more robust feature detector and descriptor are needed. Unlike SIFT, the Maximally Stable Extremal Region (MSER) detector does consider all six affine paramaters, including camera axis orientation paramaters, but the MSER feature detector normalizes these paramaters.[10] This algorithm has a limited affine invariance due to this fact and is therefore not an ideal solution in the wide-baseline problem.[3] The Affine-SIFT (ASIFT) algorithm was developed to provide the necessary robustness. The ASIFT feature detector simulates (more effective than normalizing) the camera axis orientation paramaters and uses SIFT to simulate the scale paramater and match features by normalizing for rotation and translation.[3]

3. THE PROBLEM

There currrently exists a deficiency in traditionally-collected aerial imagery datasets. The data that exists lacks a high degree of overlap, especially in oblique views. This type of imagery could be very useful in the generation of three-dimenional models, where a high density of feature matches on all surfaces in the scene are required, as well as for a variety of other uses. In order to obtain this type of imagery, it is possible to use low-cost aerial imaging systems such as remote controlled aircraft, quadcopters, and balloon-borne systems. Since the data is obtained at a low-altitude, the resulting imagery naturally has higher resolution and contains a larger degree of overlap than traditional collection methods.

Accurate geographic placement of this high-resolution, oblique data on base maps can be difficult due to the image projective geometry. Many of these systems have GPS and IMU on-board that can get the data close, but precise placement is hard to obtain due to the accuracy limitations of these devices on low-cost collection

systems. This research proposes an automatic registration technique that could make it possible to improve the positioning accuracy of the collected image data on the ground plane using existing orthorectified imagery that is available.

4. APPROACH

Once the low-altitude, high-resolution image is obtained, an initial estimate of the geographic location is provided by the on-board GPS and IMU. With this initial estimate, a bounding box is projected onto a basemap and the enclosed data is selected as a base map against which the unreferenced high resolution imagery is matched. The features matched between the two images using this method are used as tie points. The workflow implemented in this research (see Figure 1) utilizes the ASIFT algorithm to find correspondences between the two images.

To georeference an image, another image that is already georeferenced is needed as a base map. Pixels in the unreferenced image are then matched with pixels in this georeferenced image in order to relate geospatial information to those pixels in the unreferenced image. The process of matching two images requires that correspondences are established between objects in the images being matched. If a view of the same object is available in both images, then a correspondence can be obtained. However, this process is made challenging by the deformation that occurs with change in viewpoint. As the camera position changes, distortion is introduced to the image of the object, depending on the camera position.[3]

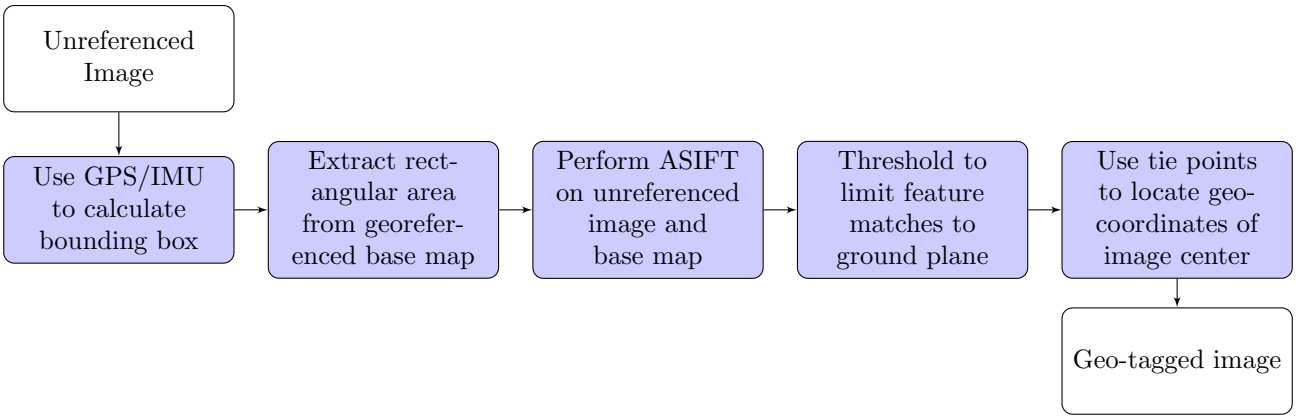
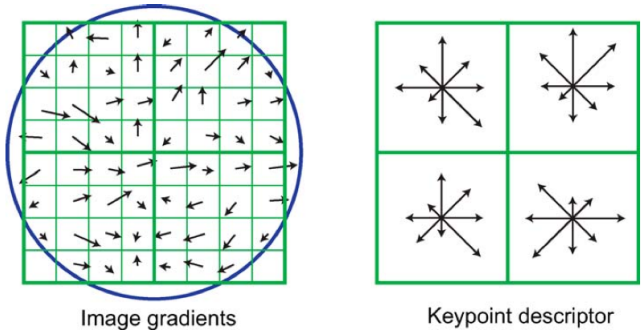


Figure 1. Proposed workflow to automatically georeference low-altitude, high-resolution oblique imagery

4.1 Feature Detection and Matching

One of the most popular approaches to finding and matching image correspondences automatically is the Scale Invariant Feature Transform (SIFT). The SIFT algorithm produces vectors that decribe distinctive invariant features in an image. These features are theoretically invariant to scale and rotation and partially invariant to change in illumination and viewpoint angle.[11] The greater the change in viewpoint, the more deformation occurs between the images. In the case of the wide-baseline problem, the change in viewpoint is large, and the Affine-SIFT (ASIFT) algorithm is required to match the images.[3]

The ASIFT algorithm simulates all possible views of a scene and then relies on the SIFT algorithm to perform the feature detection and matching.[3] The SIFT algorithm has been widely used to perform image matching and automatically find control points between images. The first step in the SIFT algorithm is to identify image feature/keypoint candidates by detecting scale-space extrema. These are locations in the image that are stable across different scales. The scale-space extrema are identified by first convolving the image with a variable scale gaussian. The scale is then incremented to create a set of Gaussian scale-space images and each set of adjacent gaussian images are subtracted to create difference of gaussian (DOG) images. The Gaussian images are then downsampled by a factor of two and the process is repeated. The extrema are located by comparing each pixel of the DOG images to its eight neighbors and the nine neighbors in the scale space above and below it. Once the extrema candidates are located, keypoints with low contrast are rejected since they are vulnerable to noise.



(a) (b)

Figure 2. The SIFT algorithm uses local image gradients to calculate a descriptor for each keypoint. (a) Image gradient magnitude and orientation are sampled around the selected neighborhood and they are weighted by a circular Gaussian window. (b) Orientation histograms are calculated over a 4x4 region (picture only shows a 2x2 region). There are eight vectors representing each bin of the histogram and the vector lengths represent the magnitude of the histogram entry.[11]

Additionally, keypoints that occur on an edge have to be eliminated, since they would be difficult to distinguish from the other keypoints on that same edge. Only the most stable keypoints remain. Next, using the direction of local image gradients, an orientation is assigned to each keypoint. Since the keypoint descriptor captures this orientation information in its description, it is this step that allows the SIFT algorithm to be rotation invariant. This keypoint descriptor is generated as the next step. The sixteen pixel neighborhood around the keypoint is used, along with the orientation assignment, to define a descriptor. To create the descriptor, image gradient magnitude and orientation are sampled around the selected neighborhood and they are weighted by a circular Gaussian window, as you can see in Figure 2(a), so that gradients farthest from the center have the least weight. Next, orientation histograms are calculated over a 4x4 region (shown as a 2x2 region in Figure 2(b)). As you can see, there are eight vectors representing each bin of the histogram and the vector lengths represent the magnitude of the histogram entry. The descriptor is generated as a vector containing each orientation histogram entry. Since there are eight bins in each histogram and a histogram for each block in the 4x4 region, the descriptor vector is 128 elements (4 x 4 x 8). The result is a vector that "describes" the keypoint. This vector is then normalized for comparison and it is this step that allows the algorithm to be partially invariant to illumination changes. These descriptors are compared to find correspondences between the images, finding points that exhibit a minimum distance between feature descriptors.[11]

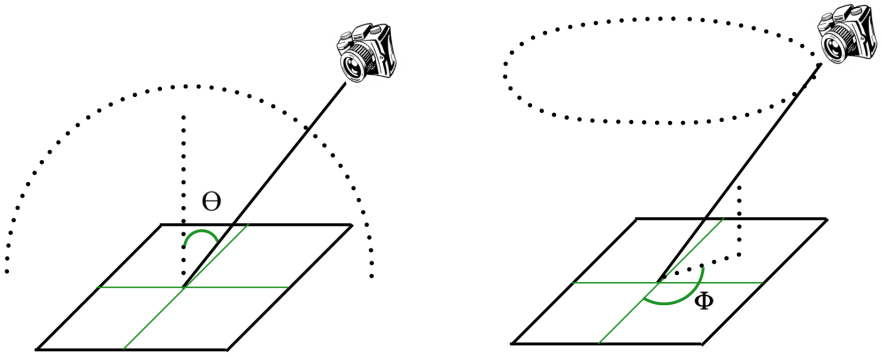
If it can be assumed that the local area of a feature can be approximated by a plane, then the deformation caused by a change in viewpoint can be approximated by an affine transform. There are six parameters that describe an affine transform. These parameters are rotation, translation, scale, and changes in the camera axis orientation. An affine transformation is represented as,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \lambda \begin{bmatrix} cos\psi & -sin\psi \\ sin\psi & cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix} \quad (1)$$

where λ corresponds to zoom, ψ represents the degree of rotation around the optical axis, $t = 1/\cos(\theta)$ where θ is the latitude angle, ϕ is the longitude angle (both latitude and longitude angles are camera axis orientation parameters), and τ is the translation.[3]

The SIFT algorithm is invariant to four of these parameters, scale, translation, and rotation.[11] The ASIFT algorithm, which is an extension of SIFT, relys on its ability to obtain those four parameters and is additionally invariant to the changes in camera-axis orientation. The ASIFT algorithm improves upon the well-known SIFT algorithm by varying the camera-axis parameters to simulate all possible views. These parameters are named latitude and longitude. According to Figure 3, the latitude parameter refers to the angle between the optical axis and a normal to the image plane and the longitude parameter refers to the angle between the optical axis

and a fixed vertical plane intersecting the optical axis.[3] The algorithm then applies SIFT to each simulated view and base map pair to find correspondences. Previously, SIFT was unable to find correspondences between wide-baseline image pairs, but the improved robustness of ASIFT makes this possible at the expense of a computationally expensive, brute force search.[3] The Maximally Stable Extremal Region (MSER) detector is suggested as a less computationally expensive approach since it attempts to normalize all six affine parameters.[10] This detector does not simulate the camera axis and scale parameters so its computation time is much lower, however, the algorithm does not succeed at being scale invariant and has a limited affine invariance.[3]



(a) Camera axis parameter 'latitude' (b) Camera axis parameter 'longitude'

Figure 3. The ASIFT algorithm simulates the two camera axis parameters, latitude and logitude.

5. RESULTS

The workflow discussed in Section 4 has been applied to sample low-altitude, high-resolution, oblique imagery provided by Pictometry International Corporation.[12] Two datasets were used, taken in New York and Nevada, with GSDs of 6 inches and 4 inches respectively. The Pictometry oblique imagery is taken at 40-45 degrees off nadir. A georeferenced base map provides a nadir view of the scene. Additionally, a low-altitude, high-resolution, oblique image is used as its wide-baseline partner. An example set of images from each dataset is provided in Figure 6. The oblique imagery contains GPS and IMU information that makes it possible to select the nadir imagery by projecting the footprint of this image on to the base map and defining a search area in the georeferenced system.

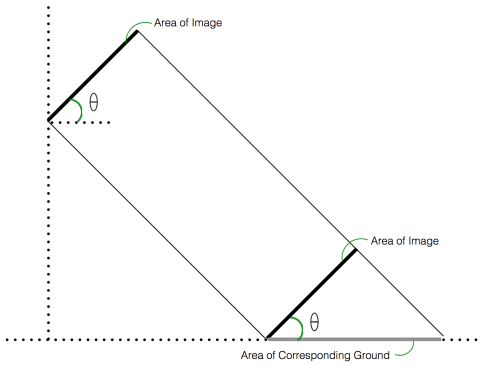


Figure 4. If the area of the image and the angle the camera is oriented are known, then the area on the ground can be calculated using similar triangles.

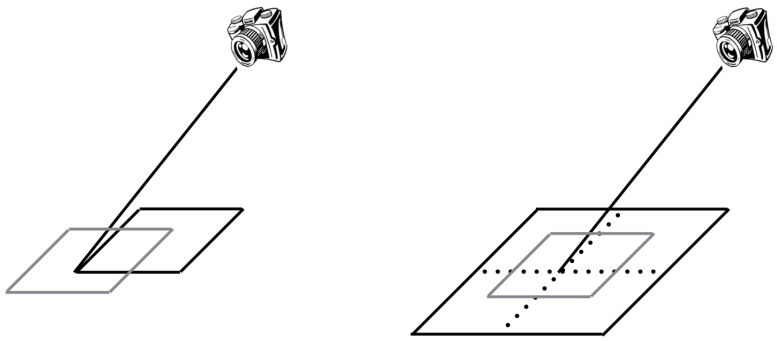
Using similar triangles, the area of the search area can be calculated with the equation,

$$GroundArea = \frac{ImageArea}{cos(\theta)} \tag{2}$$

where θ is the angle the camera is oriented off nadir (see Figure 4). Since in the pictometry system $\theta = 45\,\text{deg}$, the ground area is equivalent to,

$$GroundArea = \sqrt{2} * ImageArea \tag{3}$$

To account for any error in the GPS and IMU information, the search area will need to be roughly twice the area of the oblique image, assuming a flat earth, to ensure that the entire area is contained. Since the dataset does not provide a continuous map, the selection of the nadir image may not fall completely within one image as you can see in Figure 5(a). In this case, the four closest images will need to be stitched together which is demonstrated in Figure 5(b). The oblique image and nadir image are then run through ASIFT.



(a) (b)

Figure 5. An example situation that could occur when extracting nadir imagery from the base map imagery. (a) When selecting nadir imagery, some of the selection may fall off the edge of an image of the basemap dataset. (b) To remedy this, the four surrounding images will be stitched together before this step.

The images were matched successfully using ASIFT, which provided concept verification. As you can see in Figure 7, a majority of the correspondences are located on the ground plane. Work to further constrain points to the ground plane is necessary to obtain the most accurate registration possible. Since only a few correspondences are required to accurately place the image center, further limiting the number of correspondences matched between the images should lead to fewer keypoints located on objects other than the ground plane. A complete module is in development which will take unregistered images with GPS and IMU as input and will output an image geo-tagged with the positioning information of the center of the image.

6. CONCLUSIONS AND FUTURE WORK

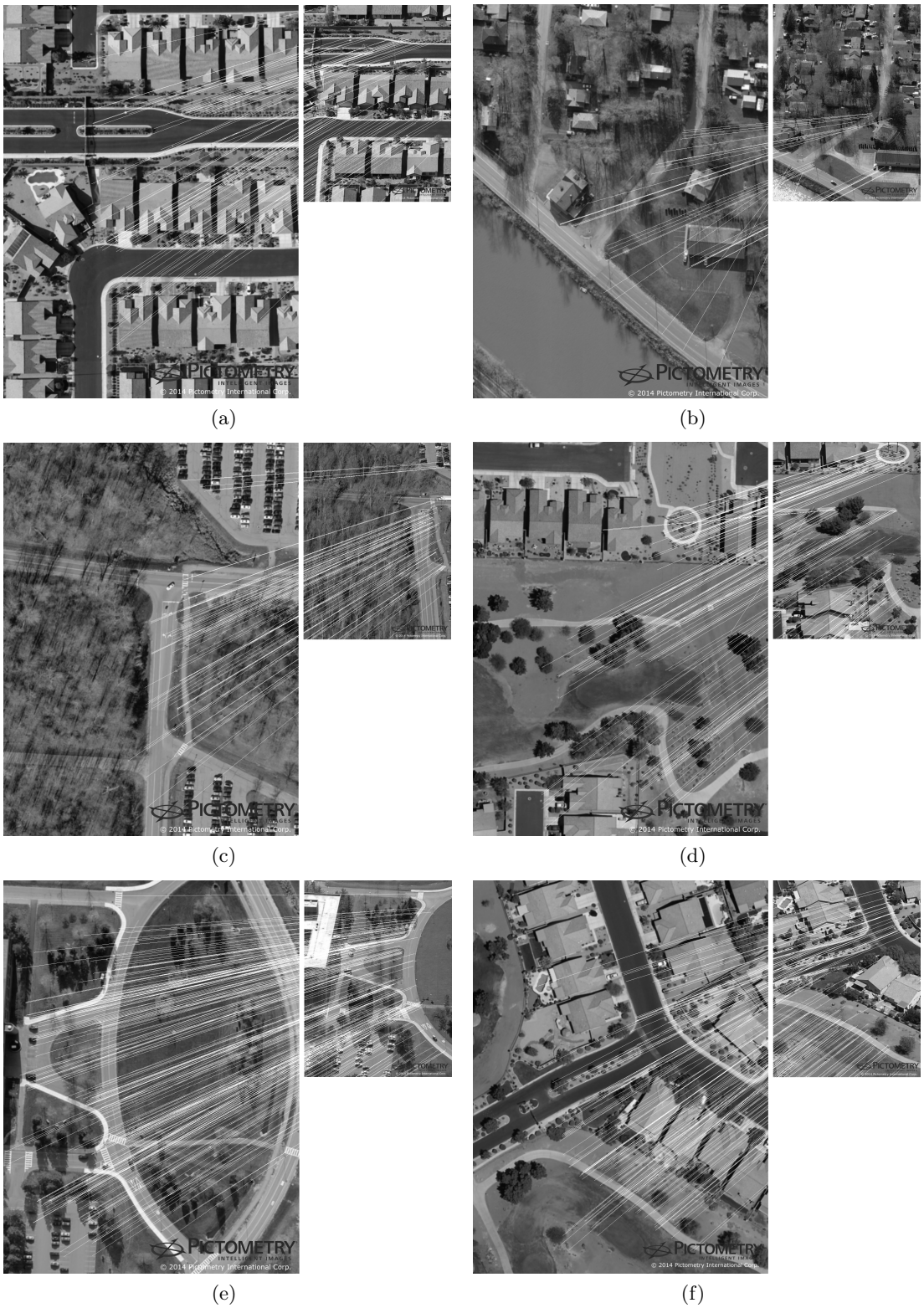
There exists a growing need for oblique low-altitude, high-resolution imagery. Since this imagery can be prohibitively expensive to collect, the use of unmanned aerial vehicles and balloon-borne systems can provide this data at a much lower cost. As this type of data collection grows, algorithms to automatically georeference this imagery will be crucial to the data's utility. The workflow presented in this paper provides an approach to the processing of these images. Preliminary results have provided the necessary concept verification.

There are additional objectives that will be adressed in future research. One of these objectives is to investigate the use of IMU data in reducing the computational time of the ASIFT algorithm. Since we know information

(a) Nadir View (b) Oblique View

(c) Nadir View (d) Oblique View

Figure 6. These images provide an example of a selection from the nadir basemap and its corresponding low-altitude, high resolution oblique image. Images (a) and (b) are taken from the Nevada dataset and (c) and (d) are from the New York dataset.



(a) (b) (c) (d) (e) (f)

Figure 7. The above examples show the success of finding correspondences using ASIFT between wide-baseline image pairs.

about the camera's positioning and pointing, it's not necessary to simulate all possible views which should reduce the run-time of the algorithm. Additionally, work to further investigate limiting correspondences to the ground plane is necessary to ensure the accuracy of the georeferencing.

ACKNOWLEDGMENTS

The authors would like to thank Pictometry International Corporation of Rochester, NY for providing the imagery used in this research as well as the National Science Foundation for funding the Advancing Innovation Research (AIR) grant under grant number IIP1127728.

REFERENCES

1. G. Verhoeven, M. Doneus, C. Briese, and F. Vermeulen, "Mapping by matching: a computer vision-based approach to fast and accurate georeferencing of archaeological aerial photographs," *Journal of Archaeological Science* **39**(7), pp. 2060–2070, 2012.
2. S. Cao, J. Jiang, G. Zhang, and Y. Yuan, "An edge-based scale-and affine-invariant algorithm for remote sensing image registration," *International Journal of Remote Sensing* **34**(7), pp. 2301–2326, 2013.
3. J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences* **2**(2), pp. 438–469, 2009.
4. J. Xiao, H. Cheng, F. Han, and H. Sawhney, "Geo-spatial aerial video processing for scene understanding and object tracking," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, June 2008.
5. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing* **21**(11), pp. 977–1000, 2003.
6. R. Berthilsson, "Affine correlation," in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, **2**, pp. 1458–1460, IEEE, 1998.
7. Q.-s. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **16**(12), pp. 1156–1168, 1994.
8. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *Medical Imaging, IEEE Transactions on* **16**(2), pp. 187–198, 1997.
9. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **27**(10), pp. 1615–1630, 2005.
10. J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing* **22**(10), pp. 761–767, 2004.
11. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), pp. 91–110, 2004.
12. "Pictometry : Intelligent images," *www.pictometry.com* , April 2014.