Ultra-wide Baseline Aerial Imagery Matching in Urban Environments

Hani Altwaijry
haltwaij@cs.ucsd.edu
Serge Belongie
sjb@cs.ucsd.edu

Department of Computer Science and Engineering
University of California, San Diego
California, USA

Figure 1: A pair of aerial images with correspondences is shown.

Correspondence matching is a fundamental problem in computer vision, having many uses in structure from motion, stereo vision, image registration, pose-estimation, and others. Today, a large amount of aerial imagery is available online via mapping services such as Google Maps. If we were to pick any pair of images from two different aerial views (see Figure 1 for an example), and perform SIFT-based [3] correspondence matching, we would find ourselves with a large number of mismatches due to the large distortions between the images. Even when augmenting these methods with typical robust approaches such as RANSAC [1] and its variants, we would still fail at finding correct correspondences since RANSAC has difficulty calculating the correct model without a large ratio of correct matches to outliers. These difficulties – large distortions, and low ratio of correct matches to outliers – together render traditional methods ineffective.

In this paper, we consider the problem of correspondence matching for aerial imagery in urban environments. Our approach builds on multiple ideas in the literature. Namely, A-SIFT [7], patch-based methods [6], Generalized RANSAC framework [8], self-similarity [5], graph-based image matching [2], and geometric-invariance [4]. The main idea behind this work is to combine view-synthesis with multiple point correspondences under a RANSAC-based scheme. Robust model estimation is supported by self-similarity principles and graph-based modeling that drives the sampling process in a restricted manner that allows the correct model to be extracted. Each of these ideas was chosen to deal with specific problems that cause failures in the earlier approaches as we will now briefly describe.

An overview of our pipeline is shown in Figure 2. We begin by extracting square patches around detected Harris corners, which we then describe using the Histogram of Oriented Gradients. Our model assumes affine distortions, which leads us to synthesizing transformations to account for probable changes between the two images. We apply our transformations to one of the input pairs only, as follows:

$$Scale = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad Shear = \begin{bmatrix} 1 & Sh_x & 0 \\ Sh_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$Rotation = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A = Scale \times Shear \times Rotation, \qquad I_{S_x,S_y,Sh_x,Sh_y,\theta} = A \times I$$

where I is an image.

Buildings, in general, exhibit a lot of repeating patterns by design which in turn leads typical methods to fail. We leverage the repeating patterns by forming a graph over self-similar patches in one of the input images. Similar patches are found by comparing their HOG descriptors. The graph is then simplified by forcing each vertex to have a single neighbor in each radial bin around it. This then yields a connected component essential for our matching step.

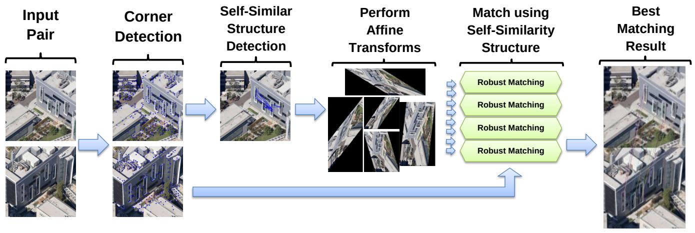

Figure 2: An overview of the matching pipeline is shown here.

Figure 3: An example pair is matched using our method, and the recovered homography is used to stitch the two images together.

During matching, patches are affine transformed and then allowed to match in a many-to-many fashion. Afterwards, our guided RANSAC approach uses the above connected component to enforce a spatial arrangement of the matches. This takes advantage of the fact that a given affine transform that yielded matches reveals more about an underlying transformation that allowed the matches to be made. We attempt to discover this by forcing spatial relationships as opposed to random sampling of probable matches, thereby increasing the probability of finding correct correspondences. A sample result is shown in Figure 3.

The main hurdles to the approach are its high computational complexity and the lack of support to multiple homographies. In our future work, we will pursue the following improvements: (1) reducing the number of affine transformations needed, (2) improving the graph operations, (3) improving the angular binning approach by including distance bins, and (4) including the support of multiple planes.

[1] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*.

[2] David Lavine, Barbara A. Lambird, and Laveen N. Kanai. Recognition of spatial point patterns. *Pattern Recognition*, 16(3):289 – 295, 1983.

[3] David G. Lowe. Object recognition from local scale-invariant features. *ICCV 1999*.

[4] Joseph L. Mundy and Andrew Zisserman, editors. *Geometric invariance in computer vision*. MIT Press, Cambridge, MA, USA, 1992.

[5] Eli Shechtman and Michal Irani. Matching local self-similarities across images and videos. In *CVPR 2007*.

[6] Saurabh Singh, Abhinav Gupta, and Alexei A. Efros. Unsupervised discovery of mid-level discriminative patches. In *ECCV 2012*.

[7] Guoshen Yu and Jean-Michel Morel. ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line 2011*.

[8] Wei Zhang and Jana Kosecka. Generalized RANSAC framework for relaxed correspondence problems. In *3DPVT 2006*.