Coarse-to-Fine Region Selection and Matching

Yanchao Yang[1,2,*], Zhaojin Lu[1,3,*] and Ganesh Sundaramoorthi[1]
[*]Joint first authors [1]KAUST, Saudi Arabia [2]University of California, Los Angeles, USA [3]Institute of Automation, Chinese Academy of Sciences

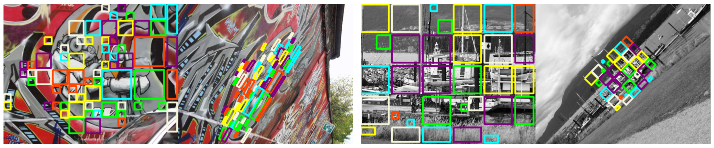

Figure 1: Visualization of the results of our algorithm on the Graffiti and Boat datasets in the Oxford dataset. Sample results of matching images with largest degree of transformation. Corresponding colors indicate corresponding regions.

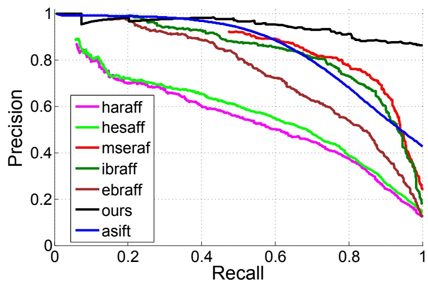

Figure 2: Precision-Recall (PR) curves for methods tested on Graffiti dataset from the Oxford dataset.

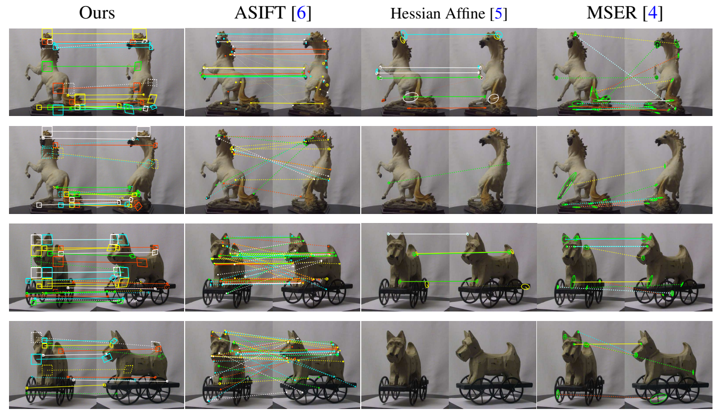

Figure 3: Sample results on the CalTech Turntable dataset. The top row of each object group are images related by viewpoint changes of 30 degrees and the bottom row is related by viewpoint changes of 60 degrees. Corresponding regions indicate corresponding regions.

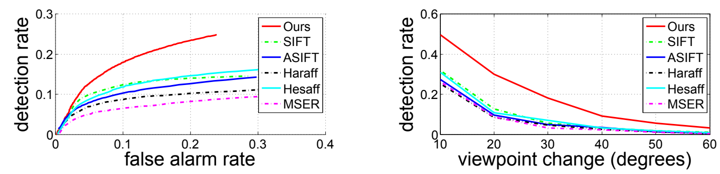

Figure 4: Quantitative results for methods tested on the CalTech Turntable dataset (used to test viewpoint). [Left]: ROC curves, and [Right]: Detection rates versus angle of rotation on the turntable for various methods tested.

In this paper, we present a new approach to wide baseline matching. We propose to use a hierarchical decomposition of the image domain and coarse-to-fine selection of regions to match. In contrast to interest point matching methods, which sample salient regions to reduce the cost of comparing all regions in two images, our method eliminates regions systematically to achieve efficiency. One advantage of our approach is that it is not restricted to covariant salient regions, which is too restrictive under large viewpoint and leads to few corresponding regions.

In our algorithm, large regions are arranged at the top of the hierarchy and smaller regions are at lower levels. The union of all regions at each level covers the entire domain. Regions are not limited to salient regions. The hierarchical decomposition enables efficient coarse-to-fine traversal through regions. Our method starts with regions at the top of the hierarchy (likely most discriminative) and proceeds to regions at lower levels of the hierarchy, which are less discriminative but more likely to match. A child region is only matched to the second image if its parent region failed to match, eliminating the need to search through all regions.

Affine transformations of regions in the hierarchy can approximate arbitrarily well any transformation arising from viewpoint change, an advantage over salient regions. Thus, we perform affine invariant matching of regions in the hierarchy. We achieve affine invariance by comparing affine orbits, which are *maximal* affine invariants. Comparison of orbits is performed by efficiently traversing the affine space. This is accomplished by a hierarchical decomposition and a coarse-to-fine search of the affine space, using ideas from [1, 2, 3]. Top levels of the hierarchy are coarse tests that eliminate large portions of the parameter space. Lower levels of the hierarchy contain tests that are increasingly selective to fewer parameters. A top-down search through the hierarchy avoids an expensive linear search.

We perform experiments to test our algorithm and compare it to other matching algorithms for wide-baseline matching. We test our method on two benchmark datasets for widebaseline, the Oxford Dataset and the CalTech turntable dataset. The Oxford dataset contains images under viewpoint change of flat scenes (leading to homographies in the imaging plane), planar rotations, and zooms. The CalTech turntable dataset concerns viewpoint change of non-planar objects, which can lead to arbitrary piecewise diffeomorphisms on the imaging plane.

Sample visualization of results on the Oxford dataset with pairs of images with the most extreme transformations are shown in Fig. 1. A sample quantitative result is shown in Fig. 2, where we have compared to standard descriptor matching methods with several different detectors. Sample visualizations on the CalTech dataset are shown in Fig. 3. Quantitative results are shown in Fig. 4. The experimental results show that our method finds more correct correspondence of the image (with fewer false alarms) than other wide baseline methods on large viewpoint change.

[1] Alexander C Berg and Jitendra Malik. Geometric blur for template matching. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–607. IEEE, 2001.

[2] F. Fleuret and D. Geman. Stationary features and cat detection. *Journal of Machine Learning Research*, 9(2549-2578):1437, 2008.

[3] Francois Fleuret and Donald Geman. Coarse-to-fine face detection. *International Journal of computer vision*, 41(1-2):85–107, 2001.

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[5] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1):63–86, 2004.

[6] J.M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

The full paper is available at the Computer Vision Foundation webpage.