## Efficient image retrieval for 3D structures

Relja Arandjelović relja@robots.ox.ac.uk Andrew Zisserman az@robots.ox.ac.uk

Department of Engineering Science, University of Oxford Parks Road Oxford, OX1 3PJ, UK

Given a query image of an object, large scale image retrieval systems usually proceed in two steps: an initial ranking/short list based on visual words [4, 6]; followed by a reranking based on a geometric relation. In order for the retrieval to be immediate both steps must be extremely efficient at run time. The visual words enable text retrieval like scalability and speed, but the reranking is potentially more costly as it involves computing a geometric relation between the query image and each target image in the short list. The geometric relation of choice has been a planar homography [2, 5].

In this work we show that epipolar geometry can be used for the ranking relation. Why is this important? As is well known [1], a planar homography is the exact relation between two images if (i) the scene is planar, or (ii) the scene is three dimensional but the camera centres of the two images coincide. In fact both these cases often apply for the type of flickr images in retrieval datasets since (i) often buildings are used for queries, and their facades are near-planar, and (ii) tourists often take photos of famous landmarks from much the same place. However, where these cases don't apply, e.g. if the situation is a 3D structure acquired from different viewpoints, potential image matches may be lost. Epipolar geometry is the exact relation for this situation.

In this paper we develop a new *minimal* solution for the affine fundamental matrix (which represents the affine approximation of epipolar geometry) based on the correspondence of a pair of ellipses. Epipolar tangency constraints on the ellipses are used to derive a quartic equation in a single variable, and this yields up to four closed form solutions for the affine fundamental matrix (Figure 2).

We evaluate the new solution using the standard framework for estimating epipolar geometry between two views: compute affine covariant elliptical regions in each image; represent these by a SIFT descriptor for matching; and then estimate the affine fundamental matrix using the RANSAC algorithm. We compare the two-ellipse solution to the four-point minimal solution [1] on a number of standard image pairs and achieve vast computational saving, with a median speedup of 8.43 and a mean of 17.48.

This solution is tailored to the architecture of large scale retrieval systems since: first, ellipses are already produced by the affine covariant region detectors (e.g. affine-Harris [3]) used by such systems; and, second, as only two correspondences are required, computation of the relation (using RANSAC) is very efficient. We show that indeed there are matches missed by the homography relation which are correctly retrieved using epipolar geometry (Figure 1), and that the system performance in terms of standard benchmarks equals or exceeds a baseline using homographies.

To our knowledge this is the first time that the ellipse pair solution has been given, and that epipolar geometry has been used to rerank a real time large scale retrieval system.

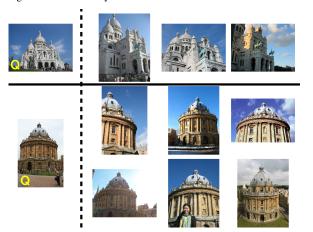
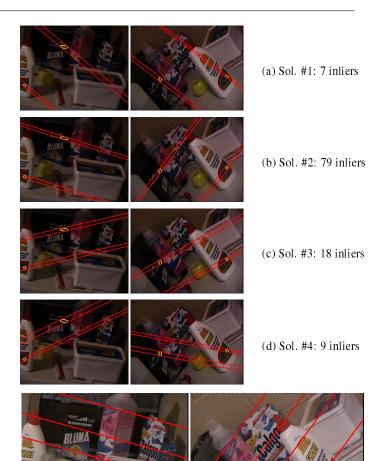


Figure 1: **Two queries from the Oxford and Paris datasets.** The results (shown right) are correctly verified with affine epipolar geometry, but are *not* verified using an affine transformation.



(e) Final solution obtained using all the inliers of solution #2: 122 inliers

Figure 2: Estimating affine epipolar geometry from two ellipses for the Wash scene. (a)-(d): the four solutions consistent with the two corresponding ellipses (shown in yellow), (b) is the correct one. Example epipolar lines are shown in red; note that they are tangent to the ellipses. (e): the final estimate. Note that there are significant perspective effects, but the method is capable of estimating a good affine approximation to the epipolar geometry.

- R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [2] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Proc. ECCV*, 2008.
- [3] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 1(60):63–86, 2004.
- [4] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proc. CVPR*, 2006.
- [5] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc.* CVPR, 2007.
- [6] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proc. ICCV*, volume 2, pages 1470– 1477, Oct 2003.