VIEWPOINT INVARIANT IMAGE RETRIEVAL FOR CONTEXT IN URBAN ENVIRONMENTS

J. P. Collomosse[1], K. Al Mosawi[1], and E. O'Neill[1]

[1] Department of Computer Science, University of Bath, Bath, Avon, U.K.

Keywords: Landmark Recognition, Skyline, Mobile, CBIR.

Abstract

We outline work in progress towards a scalable system capable of quickly identifying buildings and other outdoor urban structures from photographs. Salient characteristics of the skyline are extracted and correlated using a minimal edit distance metric to identify a subset of potential matches in a database. This subset is then refined using a slower but more precise affine invariant feature comparison to identify the best match. Our system is envisaged as an alternative to GPS where users may not have access to expensive GPS equipment, orientation may be a complicating factor (several landmarks present at a single location), or a reliable GPS fix.

1 Landmark and Skyline Recognition

Recently wide-baseline matching techniques have been applied to landmark recognition [2], but require both a calibrated camera (complicating deployment on heterogeneous clients) and a RANSAC search for each landmark (computationally expensive for large databases). We are developing a scalable two-step matching process, promising query-response times in the order of seconds over large databases, and working within the constraints of modern mobile devices.

We maintain fast lookup times over large databases by hashing images according to the salient characteristics of their skyline. Using a standard sky detector we obtain a characteristic signal by differencing points on the skyline with their best fit regression line. The signal is robust to translation, rotation and also to scale following a normalisation step. To compensate for adverse effects of skyline occlusion and foreshortening we employ a simple coding system to record salient points on the skyline. Our salience measure identifies points of high curvature that remain stable over multiple morphological scales. We encode points sequentially using a 10 symbol alphabet (Figure 1) that identifies the direction of curvature relative to the regression line; so yielding the skyline's hash-key. A default key is used when no sky is visible.

We maintain a database of urban landmarks; several training images are associated with each landmark. Each image is stored with a pre-computed hash-key. On receipt of a query image, a hash-key is generated and compared against keys in the database using a modified Levenshtein edit distance. Images with distances within a threshold are then examined using sparse affine invariant feature matching to identify the most relevant landmark. These features are computed *a priori* for each training image, and only features stable across all images for a given landmark are stored. The image recognition process runs server-side over a centralised database of landmarks. Transmitting query images over mobile networks is both slow and expensive, and we desire a solution that sends only a query's hash-key and sparse feature set. Although good results have been obtained using SIFT features, the scale space representations used to compute SIFT have significant storage overheads prohibiting use on most client devices. We are currently exploring alternative, lighter weight feature detectors based on a circular sampling technique (extending [1]).

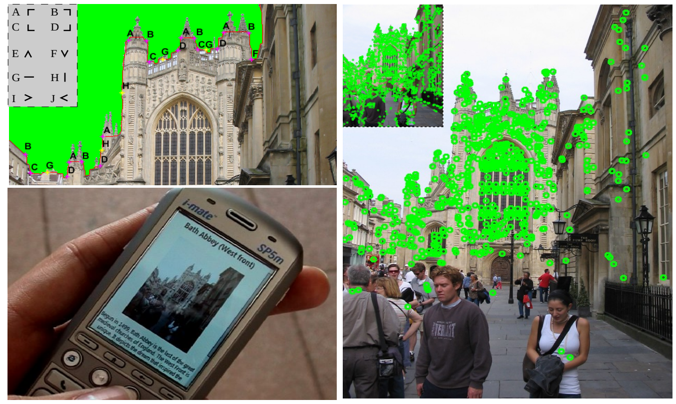

Figure 1: Illustrating our prototype. Top left: Salient points (stable over scale) encoded from the skyline and used as a hash-key. Right: Stable features identified between database training images (inset: prior to filtering). Bot. left: Mobile tourism application returns information following a successful query.

2 Early Results

We have built a demonstrator application providing information to tourists in response to photos taken at sites in Bath. Average precision and recall were 96% and 98% respectively when tested with a database of 9 sites (29 images). Initial experiments using ~ 800 stock photographs of buildings indicate our skyline hash function to scale well producing even distributions (responses remain < 5 seconds); further tests are underway. We are also working to enhance performance under adverse illumination, and exploring applications including treasure trails, virtual guided city tours, and image tagging.

References

[1] P. M. Hall, M. Owen, and J. P. Collomosse. A trainable low-level feature detector. *Proc. ICPR*, vol. 1, pp. 708–711, 2004.

[2] D. Robertsone and R. Cipolla. An image-based system for urban navigation. *Proc. BMVC*, vol. 1, pp. 260–272, 2004.