24/7 place recognition by view synthesis

Akihiko Torii[1], Relja Arandjelović[2], Josef Sivic[2], Masatoshi Okutomi[1], Tomas Pajdla[3]
[1]Tokyo Institute of Technology. [2]INRIA. [3]Czech Technical University in Prague.

We address the problem of large-scale visual place recognition for situations where the scene undergoes a major change in appearance, for example, due to illumination (day/night), change of seasons, aging, or structural modifications over time such as buildings built or destroyed. Such situations represent a major challenge for current large-scale place recognition methods. This work has the following three principal contributions.

1. First, we demonstrate that matching across large changes in the scene appearance becomes much easier when both the query image and the database image depict the scene from approximately the same viewpoint.

2. Second, based on this observation, we develop a new place recognition approach (figure 1) that combines (i) an efficient synthesis of novel views with (ii) a compact indexable image representation (VLAD encoding of dense SIFT descriptors followed by PCA compression).

3. Third, we introduce a new challenging dataset of 1,125 camera-phone query images of Tokyo [1] that contain major changes in illumination (day, sunset, night) as well as structural changes in the scene.

(a) Query image (b) Street-view (c) Synthesized view (d) Locations on the map (e) Street-view panorama (f) Associated depth-map

Figure 1: **Matching across major changes in scene appearance is easier for similar viewpoints.** The query image (a) does not match to the original database image (b) due to a major change in scene illumination combined with the change in the viewpoint. Matching a more similar synthesized view (c) is possible. The synthesized view is directly rendered from the Google street-view panorama (e) and its associated piece-wise planar depth map (f) (brightness indicates distance). Illustration of locations of (a-c) on the map (d).

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

Figure 2 shows two test query examples. The figure shows two test query images (a,d), the original street-view images (b,e), our place recognition results (c,f) (Dense VLAD descriptor with the database expanded by synthetic views) compared to the baseline method (d,g) (Sparse Fisher vectors without synthetic views [2]). Note that our method can match difficult queries with challenging illumination conditions. **Please see additional results on the project webpage [1]**

(a) Query image (b) Street-view of the query place (c) Matched synthesized view (ours) (d) Match by baseline (incorrect) (d) Query image (e) Street-view of the query place (f) Matched synthesized view (ours) (g) Match by baseline (incorrect)

Figure 2: **Example place recognition results for our method compared to baseline using only sparsely sampled feature points.** (a,d) Query image. (b,e) The original street-view image at the closest position to the query. (c,f) The best matching synthetic view by our method (correct). (d,g) The best matching street-view image by the baseline (incorrect).

[1] http://www.ok.ctrl.titech.ac.jp/~torii/project/247/.

[2] A. Torii, J. Sivic, T. Pajdla, and M. Okutomi. Visual Place Recognition with Repetitive Structures. In *CVPR*, 2013.