Speeded-up, relaxed spatial matching
Supplementary Material

Giorgos Tolias and Yannis Avrithis
National Technical University of Athens
{gtolias,iavr}@image.ntua.gr

Abstract

In this supplementary material, we present more details for the matching process, a matching example, and sample images from our Barcelona *dataset.*

Matching process

We provide here more details on erasing conflicting assignments. Function ERASE, shown in algorithm 1, assumes that two assignments are conflicting if they share the same visual word. Given a set of assignments in a bin, we first construct the set U of common visual words in the bin. This is done in line 3, where $u(c)$ is the visual word assigned to c. Then, in lines 4-5, we define for each codeword $u \in U$ the *codeword class* $e(u)$, that is, the set of all assignments mapped to u. According to our assumption all assignments in a class are (pairwise) conflicting. Therefore, we keep the strongest assignment in each class and erase the rest.

Algorithm 1 HPM ERASE

1: **procedure** ERASE(assignments C)
2: $x \leftarrow \emptyset$; $U \leftarrow \emptyset$;
3: **for all** $c \in C$ **do** $U \leftarrow U \cup u(c)$ ▷ common codewords
4: **for all** $u \in U$ **do** $e(u) \leftarrow \emptyset$ ▷ codeword classes
5: **for all** $c \in C$ **do** $e(u(c)) \leftarrow e(u(c)) \cup c$
6: **for all** $u \in U$ **do** $x \leftarrow x \cup e(u) \setminus \arg\max_{c \in e(u)} s(c)$
7: **return** erased assignments x ▷ all but strongest
8: **end procedure**

Matching example

Figure 9 illustrates matching of assignments in the *Hough voting space* for two real images. Here we present the full voting space in two 2D projections, translation (x, y) and log-scale/orientation $(\log \sigma, \theta)$, each with 16×16 bins at the fine level.

Dataset

Sample images from each of the 17 groups of the *Barcelona* dataset are presented in Figure 10. A query image from each group is presented in Figure 11 and all 5 queries for 3 selected groups are shown in Figure 12. Sample images from the set of 2M distractors used in our evaluation are shown in Figure 13. We have only collected geo-tagged images from Flickr. As a consequence, a high percentage depict urban scenery. This makes it a more challenging distractor set when used with test sets such as our own *Barcelona* dataset, *Oxford Buildings* or *Paris*.

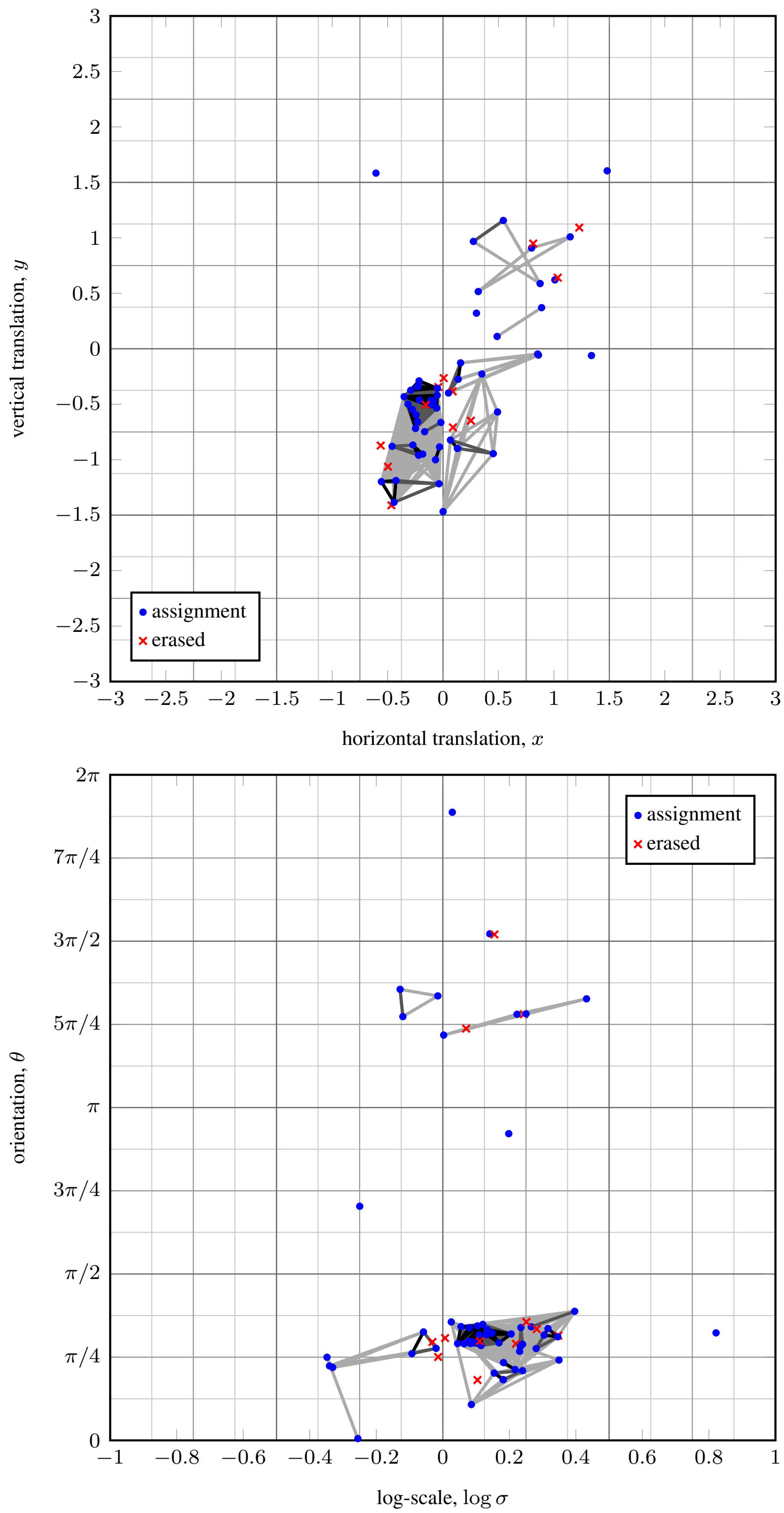

Figure 9. Correspondences as votes in *4D transformation space*. Two 2D projections are depicted, separately for translation (x, y) and log-scale / orientation $(\log\sigma, \theta)$. Translation is normalized by maximum image dimension. Maximum scale is set to 10 and orientation shifted by $5\pi/16$. There are $L = 5$ levels. Edges represent links between assignments that are grouped in levels $0, 1, 2$ only. *Level affinity* α is represented by three tones of gray with black corresponding to $\alpha(0) = 1$.

Figure 10. Representative images from all groups of the *Barcelona* dataset.

Figure 11. Selected query images from each group of the *Barcelona* dataset.

Figure 12. All query images for 3 groups of the *Barcelona* dataset.

Figure 13. Sample distractor images from the *Barcelona* dataset.