Petr Gronat

petr.gronat@inria.fr

INRIA Paris

23 av. d'Italie   
75013, Paris  
France

Dear Editor / Reviewer,

I would like to take this opportunity to explain the relationship of this manuscript to our previous and recent publications [1] “Learning and calibrating per-location classifiers for visual place recognition”, presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013 and [2] “Fisher vector places: learning compact descriptors for place recognition”, submitted to the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015. Both submissions are attached.

Most significantly we have (i) added a new section 5 that interprets the method of [2] as a special form of calibration), (ii) we have added new section on memory efficient classifier representation (Section 6), (iii) we provide additional results (Section 8) including analysis of recognition accuracy vs. compactness, new qualitative examples and (iv) we have included additional details such as pseudo code for the p-value calibration in Algorithm I and Algorithm II. The differences are explained in detail below:

**Section 1** This introduction section is an extension of both introduction sections of [1, 2].

**Section 2** This related work section mostly corresponds to Section 2 of [2]. It contains the new paragraph “Calibrating classifier scores”.

**Section 3** This section corresponds to Section 2 of [1], a new paragraph (Section 3.2) that motivates the calibrating classifier scores has been added.

**Section 4** This section extends Section 3 from [1]. The section contains the new subsection 4.3 that summarizes the calibration procedure. It also contains new pseudo codes (Algorithm I, Algorithm II) explaining the p-value calibration in offline and online stage.

**Section 5** This is anew section which interprets the method of [2] as a form of calibration. We first explain the main drawbacks of the p-value calibration procedure and motivates for usage of simpler calibration model. We take an idea of [2] and explains the method in the context of affine calibration function.

**Section 6** This is a new section that addresses efficient representation of the classifiers. It first motivates for compact classifier representation and finally it proposes an efficient representation based on dual form of the learnt SVMs. We take advantage of the fact that as only few hundreds of examples are used for training a number of support vectors is only a fraction of the dimensionality of the bag-of-words descriptor. This allows to store SVMs in a form of sparse matrix.

**Section 7** This section combines Section 4 from [1] and Secitons 4.1 and 4.2 from [2] submission to explain recent experimental setup.

**Section 8** This section presents the experimental validation of the proposed methods. Results correspond to Section 4.3 of the CVPR 2015, but for the e-SVM FV128 descriptors perform marginally better as we have found better regularization parameters parameters.

**- Section 8.3** This new section provides an analysis of recognition vs. performance for proposed methods and shows the benefits of the affine calibration (w-norm) on learnt Fisher vectors.

**Figure 9** This new figure shows a recognition performance vs. memory requirements for Pittsburgh 25k dataset. Comparing to figure 3 from [2], that analyze a memory complexity vs. Fisher vector dimension, the new figure shows a memory requirement on the y-axis and it also shows a memory complexity for other methods (BOW, BOW p-val, BOW w-norm).

**Figure 4** is a new figure showing a recall curves for bag-of-visual-words baseline and two different calibration methods (p-val and w-norm). The figure demonstrates that both methods improve the recognition performance over the baseline for different lengths of a shortlist.

**Figure 6** A new figure with examples with correctly and incorrectly localized queries learnt for the bag-of-visual-word representation.

**Figure 8** A new figure with examples with correctly and incorrectly localized queries learnt for the Fisher vector representation.

Kind regards

Petr Gronat