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 publication [1] “Learning and calibrating per-location classifiers for visual place recognition”. Most significantly we have (i) added a new section 5 that describes a special form affine calibration that does not require construction of empirical cumulative density function, (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].

**Section 2** This is a new section addressing related work which addresses state-of-the-art methods for place recognition, per-exemplar support vector machines and classifier score calibration methods.

**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 of [1]. This section contains a new subsection 4.3 that summarizes the calibration procedure. It also contains new pseudo codes (Algorithm I, Algorithm II) describing the p-value calibration in offline and online stage.

**Section 5** This is anew section that addresses complexity of p-value calibration and proposes different approach for classifier score calibration. We first explain the main drawbacks of the p-value calibration procedure in terms of complexity and than motivate for usage of simpler calibration model based on affine calibration function. A new calibration method is proposed based on analysis of the per-exemplar SVM objective 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** The section extends Section 4 of [1] and explains recent experimental setup.

**Section 8** This section presents the experimental validation of the proposed methods.

**- 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 4** This new figure illustrates an effect of different regularization on learnt hyperplane for per-exemplar SVM and demonstrates when learnt hyperplane can be interpreted as a new descriptor.

**Figure 5** 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 7** A new figure with examples with correctly and incorrectly localized queries learnt for the bag-of-visual-word representation.

**Figure 9** A new figure with examples with correctly and incorrectly localized queries learnt for the Fisher vector representation.

**Figure 10** This new figure shows a recognition performance vs. memory requirements for Pittsburgh 25k dataset. The plot show recognition accuracy (y-axis) as a function of memory requirement of proposed methods (x-axis).

Kind regards

Petr Gronat