NEIGHBORHOOD REVERSIBILITY VERIFYING FOR IMAGE SEARCH

Zhong Zheng, Yao Zhao, Shikui Wei, Zhenfeng Zhu

Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, China
Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, 100044, China
shkwei@bjtu.edu.cn

ABSTRACT

The neighborhood structure can significantly impact the effectiveness of image search, and fulfilling the reversibility of neighborhood may improve the image search quality. This paper proposes an effective and efficient scheme for reconstructing the symmetry relationship of *k*-nearest neighborhood (KNN). In particular, we design a verifying function to learn the prior knowledge of neighborhood reversibility among images. By exploiting the prior knowledge, the image search system will give higher rank to those images that satisfy the reversibility of KNN relationship with the query. In addition, we systematically investigate the sensitivity of neighborhood size on image search quality and propose an adaptive selection scheme for improving robustness of neighborhood reversibility learning methods. The extensive experimental results show that the proposed scheme remarkably improves the image search quality and give a comparable but more stable performance to the state-of-the-art method for various image datasets.

***Index Terms*—**Neighborhood, Reversibility, Image, Search, Cross-Media

1. INTRODUCTION

Image search is an important but still challenging research area, which addresses the problem of searching for images similar to the query. In recent years, lots of works have been done for handling some challenging issues such as the appearance variations in illumination, scale, viewpoint, orientation [1, 2]. As a kind of effective methods, local descriptors based schemes are designed on the basis of local descriptor matching of images [3]. However, while this kind of methods is quite effective for handling some difficult appearance variations due to their robustness, it leads to high computational complexity. To simplify the search process, more recent works employ the bag-of-words (BoW) framework [4] to facilitate large-scale image search, in which local descriptors are individually quantized to visual words against a learned visual vocabulary. Since visual words are only approximate representation of local descriptors, BoW-based framework will inevitably decrease the image search quality due to the quantization errors. Therefore, some works have been done to address the issue. One kind of methods is to introduce complementary information such as binary embedding code [5] to supplement information loss caused by quantization. An alternative approach is to employ soft-assignment strategy [6] or spatial information to alleviate the effect of quantization errors. For example, query expansion with spatial information is reported in [7-10], and weak geometric consistency [5] is exploited for reranking the initial search list via spatial verification.

In addition to image representation, many works also pay attention to similarity measurement among images, such as Euclidean distance, cosine distance [4], TF-IDF scoring [11]. In [12], a new distance measure was proposed by exploiting probabilistic relationship between visual words so as to improve matching quality. Furthermore, descriptor projection learning [13] and hierarchical methods [11] are reported for improving the effectiveness and efficiency of image matching. However, while lots of similarity measurement methods are reported in previous works, most of them take only unidirectional neighborhood relationship into account when calculating the similarity between two images. That is, only the neighborhood structure from query image to database images is considered, i.e., neighborhood non-reversibility. For example, if image A is one of B's KNN, it is possible that B is not A's KNN, which results in unbalanced neighborhood measure and false matches [14, 15].

In this paper, we focus mainly on the reversibility of neighborhood relationships by explicitly learning the reversibility property among images. We propose an effective and efficient scheme for reconstructing the reversibility relationship of KNN. With the designed reversibility verifying function, the prior knowledge of neighborhood reversibility among images is learned in an offline manner and then is employed to improve the initial search results obtained by using any distance measures. In addition, the effect of neighborhood size is fully investigated on various image datasets, and an adaptive selection scheme of neighborhood size is proposed to improve the robustness of reversibility verifying process. Experimental results show that the proposed scheme can remarkably improve the whole image search quality.

2. PRELIMINARIES

The following introduces the preliminary techniques of image search and analysis of reversibility relationship, which include three main components: (1) Build an initial image search system based on the Bag-of-Words framework; (2) Introduce and analyze the key observation on reversibility; (3) Analyze the sensitivity of neighborhood size. Each component will be detailed in the following subsections.

2.1 BoW-based image search framework

Before we discuss the reversibility of the neighborhood, we need to construct an initial image search system based on some standard distances. In this work, we construct an image search system based on the BoW framework. The key idea is to learn a visual vocabulary by quantizing the training local descriptors. Then each image is represented by a visual word vector. Employing some standard distance measure, we can perform image search process by calculating the distances between the visual word vector of query image and the visual word vectors of database images. In our experiments, the hierarchical k-means [11] algorithm is used to learn the visual vocabulary, where two-level structure is employed and 100K visual words are learned. For distance measure, we employ the commonly used cosine distance [4].

2.2 Reversibility of neighborhood relationship

The real-world image search process shows that some top-ranked results that are of unidirectional neighborhood relationship with query are generally irrelevant, while the returned images with bidirectional neighborhood relationship are true relevant with query. This observation was also reported in [15].

As argued in [15], it is possible to improve the accuracy of image search by modifying the neighborhood structure. To this end, a contextual dissimilarity measure (CDM) is reported in [14, 15] to improve the symmetry of the neighborhood by updating the distances such that the average distance of a vector to its neighborhood is almost constant, which significantly improves the whole image search quality. As an alternative, the reversibility information can also be involved into the similarity calculation process by explicitly verifying the reversibility property between query and each returned image. The verifying process is performed by conducting additional search process using the returned images as queries.

2.3 Sensitivity analysis of the neighborhood size

As discussed above, the prior knowledge of neighborhood reversibility can be involved into the image search process by two strategies, i.e., improving the symmetry of KNN relationships or explicitly verifying reversibility. For the first strategy, we need to set the number of nearest neighbors for each image before updating distance matrix [14, 15]. Likewise, the length of returned short list (i.e, the number of nearest neighbors) for each image should also be set before verifying the reversibility of neighborhood. That is, the neighborhood size is a key factor that can affect the image search quality for both strategies. However, less effort has been done for systematically investigating the sensitivity of neighborhood size to neighborhood reversibility learning, especially on various image datasets. For example, CDM algorithm in [14, 15] only analyzes the effect of the neighborhood size k on the N-S dataset, which is only specific for the dataset. In order to discover some general rules, we carefully investigate the effect of neighborhood size on different neighborhood reversibility learning strategies with various image datasets. Here, two commonly used image datasets are employed, i.e., Holidays dataset [5] and Oxford dataset [9]. As to be shown in experimental part, the neighborhood size k indeed has a great impact on the performance of the algorithms, and the optimal size is quite different with various image datasets.

After carefully analyzing two image datasets, we find that the main difference lies in the average number of true relevant images of queries. There are average 2.98 true relevant images for each query in the Holidays dataset, while the Oxford dataset corresponds to average 51.4 true relevant images for each query. This observation indicates that the best k is probably related to the number of relevant images. As to be shown in experiment part, the optimal k should be big enough so that most of true relevant neighbors can be involved to learn the neighborhood reversibility. This motivates us to develop an adaptive strategy for choosing the neighborhood size k so as to handle various image datasets.

3. ADAPTIVE REVERSIBILITY VERIFYING

As stated in [15], explicitly verifying reversibility requires to conduct more additional image search processes, i.e., per returned image a short list, leading to more computational cost and memory usage at the online stage. In this section, we propose two reversibility verifying schemes, i.e., hard verifying and soft verifying, which can smartly avoid performing additional queries by storing a single value for each image at the offline stage. In addition, we also propose an adaptive strategy to automatically select the neighborhood size k for each image so as to improve the robustness of neighborhood reversibility learning methods.

3.1 Hard reversibility verifying

If A is one of B's KNNs and B is one of A's KNNs, we call that A and B are KNN-reversible. The key idea of reversibility verifying is to directly verify if query and returned images are KNN-reversible. The ones that are KNN-reversible with query are directly put into the top of the returned list. In order to judge if one returned image is KNN-reversible with the query, we store the k-th distance values for each image at offline preprocessing stage and then employ the distance to facilitate reversibility verifying. This method includes offline preprocessing stage and online search stage as follows:

Offline preprocessing stage: Each image in dataset is treated as a query to perform a search process across the whole dataset, and the distance between the image and its k-th nearest neighbor is calculated and stored. The k-th distance T_i of any image w_i can be denoted as follows:

$$T_i = D(w_i, w_{ik}), \tag{1}$$

where w_{ik} denotes the k-th nearest neighbor of w_i. $D(. , .)$ denotes the distance measure of two images. Compared to the visual word vectors and inverted file of vocabulary, the storage overhead for k-th distance values is negligible.

Online search stage: After performing a search process for any query q, we can obtain the distance values between the query and its KNN neighbors. If the distance between q and one of its KNN neighbors w_i is lower than T_i, q and w_i must be KNN-reversible with each other. Then the image w_i will be directly put into the top of result list. In this way, much more additional search processes are avoided.

3.2 Soft reversibility verifying

For the hard reversibility verifying scheme, top-k returned images (i.e., KNNs) are first extracted from the initial returned list and then reranked according to the reversibility verifying information. After that, a new result list is built by substituting the top-k returned images in initial list with the reranked ones. That is, those images that don't meet the KNN-reversible property are still retained in the top-k results, while they may be put to the rear of the k-length list. In this way, the improvement of image search quality is limited, which heavily depends on the number of true relevant images of query. For example, if the number of true relevant images is far higher than k, it is more possible that most of top-k images are true relevant images and their proportion is high. In this case, the hard scheme will fail since search quality changes nothing no matter how you rerank the list. To address this issue, we propose a soft reversibility verifying scheme. Instead of reranking only the top-k returned results, the proposed scheme can rerank the whole initial returned list by assigning a reversibility weight to each returned image via the proposed reversibility verifying function. This scheme also includes offline and online stages.

Offline preprocessing stage: Similar to hard reversibility verifying, we need to calculate and store a k-th distance value for each image. Therefore, this stage is the same with the hard reversibility verifying scheme. Our main contribution lies in the online stage.

Online search stage: After the initial result list for the query q is returned by some standard distance measure, we calculate a reversibility weight for each returned image w_i using the following reversibility verifying function:

$$f(q, w_i, T_i) = \left(\frac{D(q, w_i)}{T_i} \right)^2, \tag{2}$$

where $D(q, w_i)$ is the distance between query q and image w_i, T_i is k-th distance obtained for image w_i at offline stage. $f(q,w_i,T_i)$ conveys the degree of reversibility. If $f(q,w_i,T_i)$ is lower than 1, it means that q and w_i are KNN-reversible with high probability, and vice versa.

By introducing the reversibility weight into image search process, we can update the original distance measure as follows:

$$D^*(q, w_i) = D(q, w_i) \times f(q, w_i, T_i), \tag{3}$$

where $D^*(q,w_i)$ is the updated distance measure. As indicated in the equation, if q and w_i are possibly KNN-reversible, $f(q,w_i,T_i)$ will make the original distance much smaller, resulting in higher rank for w_i in the returned lists. Likewise, the irrelevant images in the top-k list can be moved out since their updated distances are bigger than their original ones. Experimental results confirm that soft scheme can result in better performance than the hard one.

The difference between our schemes and the strategy of verifying the reversible property mentioned in [15] is that our scheme verifies the reversible property by the prior knowledge acquired at the offline stage. Hence, little computation is added to the online search procedure.

3.3 Adaptive selection of neighborhood size

In Section 2.3, we have systematically discussed the sensitivity of neighborhood size k on learning the reversible property. The optimal neighborhood size k is quite different for various datasets. To alleviate the effect of sensitivity and provide a more stable performance improvement, we design an adaptive strategy to automatically select the appropriately optimal neighborhood size k for each image at the offline preprocessing stage.

Instead of directly setting a fixed neighborhood size k for all images, we separately select a k_i and calculate k_i-th distance for a specific returned image w_i. Given an initial returned result list when treating w_i as query, we firstly calculate a mean distance D_{im} of top-M results in the list as follows:

$$D_{im} = \frac{1}{M} \sum_{w_x \in W_{iM}} D(w_i, w_x), \tag{4}$$

where W_{iM} denotes the M nearest neighbors of image w_i in the dataset. D_{im} conveys the approximate distance between w_i and its true relevant images. In fact, it is based on an underlying assumption that a few top-ranked results are true relevant to the query. The assumption is commonly used in pseudo-relevance feedback algorithms [1].

Based on the approximate distance D_{im}, we define a difference function for any rank position k_x in the initial result list:

$$P_{ix} = \frac{D(w_i, w_{ik_x}) - D_{im}}{D(w_i, w_{ik_1}) - D_{im}}, \tag{5}$$

where w_{ik_x} denotes image w_i's k_x-th nearest neighbor, and $D(w_i, w_{ik_x})$ is the distance between w_i and its k_x-th nearest neighbor. Note that k_1 is a benchmark value which is not more than k_x. To simplify the computation, we limit k_x to the set $\{k_1, \ldots k_x, \ldots, k_n\}$, in which k_x is greater than k_{x-1}, and k_1 is the benchmark value.

As indicated, the value of P_{ix} becomes bigger as k_x is increasing. Obviously the probability that w_{ik_x} is a true relevant image of w_i becomes lower when P_{ix} is getting bigger. Therefore, an optimal k_x should be the biggest value that can guarantee a high probability that w_{ik_x} is true relevant image of w_i. To this end, a threshold value ε is set to select the optimal k_x. In our scheme, if P_{ix} is going up to ε at a certain k_x when increasing k_x from k_1 to k_n, k_{x-1} is selected as the optimal neighborhood size of w_i. The value of ε is selected empirically in our experiments, which is not sensitive to various image datasets.

4. EXPERIMENTS

4.1 Experimental setup

We use the Oxford Buildings dataset available from [16] and Holidays dataset available from [17] as our test datasets to evaluate our scheme. The Oxford dataset consists of 5062 images and 55 query images. The Holidays dataset consists of 1491 photos and 500 query photos of different objects and landscapes. For both image datasets, the SIFT features [3] are extracted with the Hessian-affine detector for each image. In addition, a subset from Flickr60K that is available in [17] is employed to train our visual word vocabulary. In our experiment, the size of vocabulary is fixed to 100K. In fact, similar conclusion can be drawn with other sizes of vocabularies.

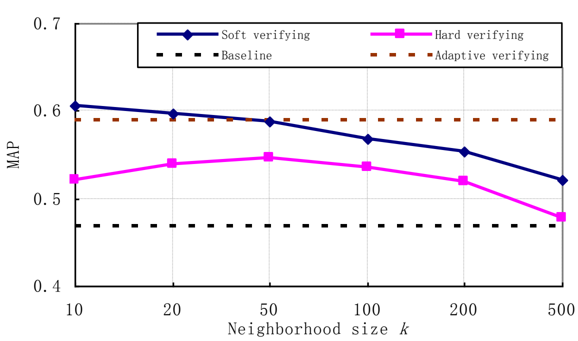

Fig. 1. Comparison of various verifying schemes on Holidays dataset.

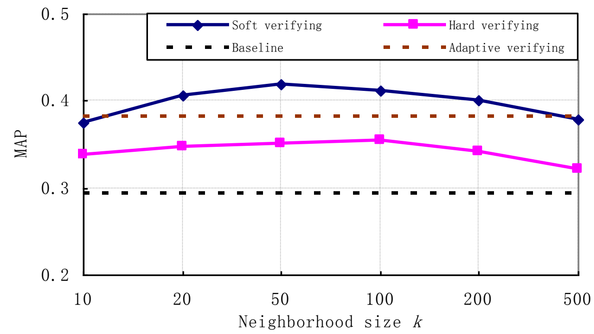

Fig. 2. Comparison of various verifying schemes on Oxford dataset.

For evaluation criteria, the mean average precision (MAP) is employed to evaluate the image search quality as done in [6, 9], which can evaluate the overall search performance of multiple queries.

4.2 Evaluation on various verifying schemes

In this subsection, various verifying schemes are tested on two very different image datasets and various neighborhood sizes. The experimental results are demonstrated in Figure 1 and 2. Note that the performances of both baseline and adaptive verifying scheme keep constant since no neighborhood size needs to be selected for these two schemes. Here, we plot them at the same figure to facilitate the comparison. We can clearly observe that all the verifying schemes significantly outperform the baseline. This means that image search could benefit from the reversibility information of neighborhood. As expected, both soft reversibility verifying schemes and adaptive verifying scheme are remarkably better than the hard one on both image datasets. That is, the soft verifying scheme indeed addresses the problem in hard verifying scheme. In addition, we can also observe that the adaptive scheme achieves the best or comparable performance to the soft verifying scheme. This means that our proposed adaptive strategy is very effective for automatically selecting optimal neighborhood size. To intuitively illustrate the improvement, we provide two search examples on both Holidays and Oxford datasets respectively. The results are illustrated in Figure 3. The first line shows the results returned by the baseline. The second line shows the results returned by applying the soft reversibility verifying scheme. It is clear that the soft reversibility verifying scheme successfully removes some irrelevant images and adds more relevant images to the top-ranked results.

4.3 Sensibility evaluation

Fig. 3. Illustration of search examples before and after introducing soft verifying scheme.

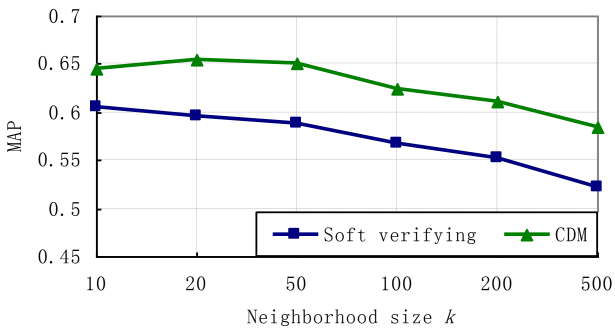

Fig. 4. Impact of *k* on Holidays dataset.

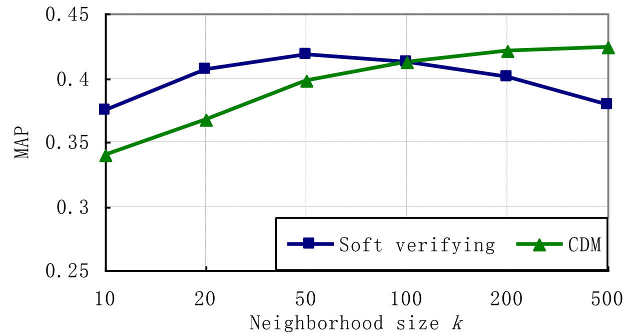

Fig. 5. Impact of *k* on Oxford dataset.

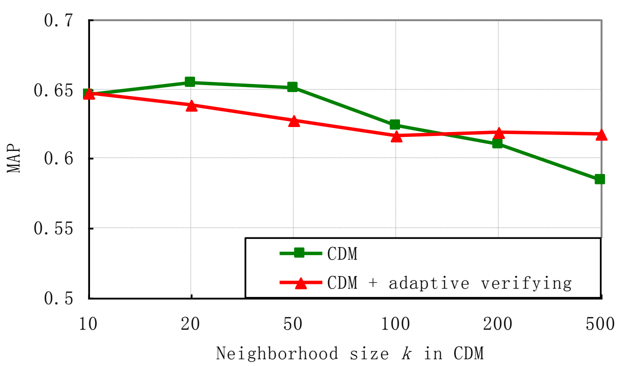

Fig. 6. Comparison of CDM and CDM+adaptive verifying scheme on Holidays dataset.

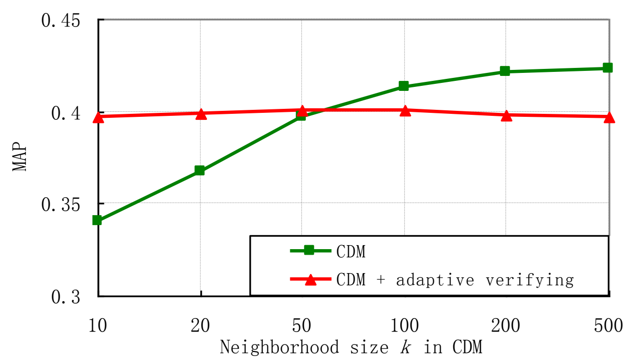

Fig. 7. Comparison of CDM and CDM+adaptive verifying scheme on Oxford dataset.

In this subsection, we conduct several experiments to evaluate the sensibility of neighborhood size on different neighborhood reversibility learning schemes. Since the hard reversibility verifying scheme is far inferior to both soft one and other existing schemes, we only test the soft reversibility verifying scheme and a state-of-the-art scheme (i.e., CDM). Figure 4 and 5 demonstrate the experimental results. As shown, the neighborhood size k has a significant effect on the performance of image search, and the optimal k even for the same scheme is quite different on various image datasets. For example, CDM gives its best performance on Holidays dataset when neighborhood size k is set to 20, while 500 is the optimal one on Oxford dataset. In brief, both neighborhood reversibility learning schemes are sensitive to neighborhood size, and there is no fixed k that can make neighborhood reversibility learning schemes work well on various image datasets.

4.4 Comparison of various methods

As stated above, CDM and reversibility verifying schemes are two kinds of solutions for learning neighborhood reversibility, both of them are sensitive to the neighborhood size. By introducing the adaptive strategy to reversibility verifying, we can greatly alleviate the issue. In this subsection, we attempt to combine the two kinds of solutions and provide a comparable and more stable performance to the state-of-the-art methods. In our experiments, CDM scheme is combined with the proposed adaptive reversibility verifying scheme, in which the outputted result list from CDM scheme is treated as the input of adaptive reversibility verifying scheme. The experimental results on two image datasets are illustrated in Figure 6 and 7 respectively. As shown, after introducing the proposed adaptive scheme, the performance of CDM is comparable but more stable to the original CDM even with various neighborhood sizes. For the Holidays dataset, the MAP always slightly changes between 0.6 and 0.65 no matter which k is selected for CDM. For the Oxford dataset, the MAP almost stays around 0.4. This means that the proposed adaptive scheme indeed alleviates the sensibility of neighborhood size on CDM scheme.

5. CONCLUSION

This paper introduces a simple but effective scheme to learn the reversibility information of neighborhood for improving the image search quality. In particular, a reversibility verifying function is designed to calculate the reversibility weight, and then a soft reversibility verifying scheme is proposed by involving the weight to the verifying process. In addition, an adaptive strategy is proposed to deal with the problem of sensitivity on neighborhood size. The experimental results show that introducing the proposed scheme to image search system can significantly improve image search quality. And combining the adaptive scheme with the state-of-the-art CDM approach can greatly alleviate the sensitivity of neighborhood size of CDM and provide more stable but comparable performance on various image datasets.

6. ACKNOWLEDGEMENTS

This work was supported in part by the 973 Program (No. 2012CB316400), PCSIRT (No.IRT201206), the National Science Foundation of China (No.61202241, No.61210006, and No.61025013), the Fundamental Research Funds for the Central Universities (No.2013JBM024), and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

7. REFERENCES

[1] S. Wei, Y. Zhao, Z. Zhu, and N. Liu, "Multimodal fusion for video search reranking," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 8, pp. 1191-1199, 2010.

[2] S. Wei, Y. Zhao, C. Zhu, C. Xu, and Z. Zhu, "Frame fusion for video copy detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 1, pp. 15-28, 2011.

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[4] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision*, 2003.

[5] H. Jégou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *European conference on computer vision*, 2008.

[6] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[7] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in *IEEE International Conference on Computer Vision*, 2007.

[8] O. Chum, A. Mikulik, M. Perdoch, and J. Matas, "Total recall II: Query expansion revisited," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[9] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[10] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[11] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[12] A. Mikulik, M. Perdoch, O. Chum, and J. Matas, "Learning a fine vocabulary," in *European Conference on Computer Vision*, 2010.

[13] J. Philbin, M. Isard, J. Sivic, and A. Zisserman, "Descriptor learning for efficient retrieval," *European Conference on Computer Vision*, 2010.

[14] H. Jégou, H. Harzallah, and C. Schmid, "A contextual dissimilarity measure for accurate and efficient image search," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[15] H. Jégou, C. Schmid, H. Harzallah, and J. Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no.1, pp. 2-11, 2010.

[16] http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/

[17] http://lear.inrialpes.fr/~jegou/data.php