# QUERY EXPANSION WITH DIFFUSION ON MUTUAL RANK GRAPHS

*Xiaomeng Wu, Go Irie, Kaoru Hiramatsu, and Kunio Kashino*

NTT Communication Science Laboratories, NTT Corporation

**ABSTRACT**

In query expansion for object retrieval, there is substantial danger of query drift, where irrelevant information is inferred from pseudo-relevant images to enrich the query. To address this issue, we propose a query expansion method from the viewpoint of diffusion. It explores the structure of highly ranked images in a topological space, assuming that false positives reside on different manifolds from the query. For this purpose, a mutual rank graph is defined on pseudo-relevant images, and their distribution is learned by diffusing their query similarities through the graph. The relevance of a database image can thus be obtained by marginalizing over the learned distribution. The mutual rank graph accounts for varying local density in the image space, leading to great robustness as regards query drift and high generalization ability. The proposed method experimentally shows a consistent boost in the performance of object retrieval with handcrafted features on standard benchmarks.

***Index Terms*** — Diffusion, object retrieval, query expansion

## 1. INTRODUCTION

Object retrieval is an important tool that supports various applications, e.g., content-based image browsing [1], visual localization [2], and 3D reconstruction [3, 4]. One problem with object retrieval relates to changes beyond built-in feature invariance, e.g., large illumination and viewpoint change, occlusion, etc., and the limited visual information available from the query. One way to compensate for this deficiency is query expansion (QE) [5–12], which assumes that most highly ranked images are relevant and have a sufficient number of variations. The coverage of visual aspects can then be improved by reissuing highly ranked images as an expanded query.

The above assumption is not always correct due to the imperfect discriminative power of image representations. False positives among high-ranked images may make the inference of the expanded query diverge from the depicted object, which is known as query drift [5]. To address this issue, spatial verification has been widely used to discard geometrically inconsistent image features [5, 11, 12]. Tolias and Jégou [12] also imposed some similarity and frequency constraints on image features to pursue a higher precision of pseudo-relevant images. These studies avoid query drift at the feature level and so are limited to specific image representations. Qin et al. [8] developed criteria based on reciprocal nearest neighbors to decide whether to use an image for QE. The method is parameter-sensitive and suffers from asymmetric primary ranking biased by an arbitrary cutoff [10]. Shen et al. [11] proposed the spatially-constrained similarity measure (SCSM), which weights pseudo-relevant images on the basis of their positions in the primary ranking. Arandjelovic and Zisserman [9] proposed discriminative QE (DQE), which trains a linear SVM with high and low-ranked images. All database images are then ranked by their distance from the decision boundary. Nevertheless, false positives may lead to poor hyperplane fitting when they dominate a set of pseudo-relevant images.

We focus on the avoidance of query drift from a different viewpoint. Our objective is to learn a probability distribution of images, which are highly ranked in a primary search and referred to as pivots hereafter, in terms of their query relevance. The probability that a database image is positive can be obtained by marginalizing over the learned distribution. Observing pivots, we find that true positives are usually on the same structure, e.g., either a cluster or a manifold, with high similarity while false positives are on different structures. We also assume that the similarity between the relevances of two pivots should be consistent with their visual similarity. On the basis of these assumptions, we propose learning the distribution of pivots with diffusion [13–17]. A mutual rank graph, which accounts for varying local density in the image space, is defined on the pivots. The pivot distribution is learned by diffusing their query similarities through the graph. Unlike previous studies [5–12], our method explores the structure of pivots in a topological space and yields greater robustness as regards query drift when false positives reside on different manifolds from the query. Against standard benchmarks, our method consistently outperforms the state of the art techniques based on handcrafted features with low parameter sensitivity.

Studies on shape retrieval [18–20] and Iscen et al.'s method [21] apply diffusion globally to all database images. In comparison, our method constrains diffusion semi-locally by the nearest neighbors of the query, i.e., the pivots, and so avoids the potential coexistence of different classes on a global manifold. This constraint also avoids the need for an online overhead for updating a huge graph each time an unseen query comes in. Our method is very efficient and provides a practical query time of less than one second per 100K images.

## 2. PROPOSED METHOD

### 2.1. Query Expansion

Given a query $\mathbf{q}$ and a dataset $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the goal of object retrieval is to rank all the images in $\mathcal{X}$ by their probability $p(\mathbf{x}|\mathbf{q})$ of having a common object with $\mathbf{q}$. Hereafter, we replace the notation $p(\mathbf{x}|\mathbf{q})$ with $p(\mathbf{x})$ for simplicity.

Let $s(\cdot, \cdot)$ be a symmetric and positive similarity measure. We denote images that are highly ranked by $s(\mathbf{q}, \mathbf{x})$ in a primary search as $\mathbf{z}_1, \ldots, \mathbf{z}_m$. We then have a set of pivots $\mathcal{Z} = \{\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_m\}$ with $\mathbf{z}_0 = \mathbf{q}$ and $m \ll n$. The conditional probability of $\mathbf{x}_i \in \mathcal{X}$ given $\mathbf{z}_j \in \mathcal{Z}$ can be defined by

$$p(\mathbf{x}_i|\mathbf{z}_j) = \frac{s(\mathbf{x}_i, \mathbf{z}_j)}{\sum_{i'=1}^{n} s(\mathbf{x}_{i'}, \mathbf{z}_j)}. \tag{1}$$

Suppose that we have obtained a distribution $p(\mathbf{z})$ over $\mathcal{Z}$ indicating the probability that $\mathbf{z}$ is relevant to $\mathbf{q}$. We can obtain the probability $p(\mathbf{x})$ by marginalizing over $p(\mathbf{z})$:

$$p(\mathbf{x}_i) = \sum_{j=0}^{m} p(\mathbf{x}_i, \mathbf{z}_j) = \sum_{j=0}^{m} p(\mathbf{x}_i|\mathbf{z}_j)p(\mathbf{z}_j). \tag{2}$$

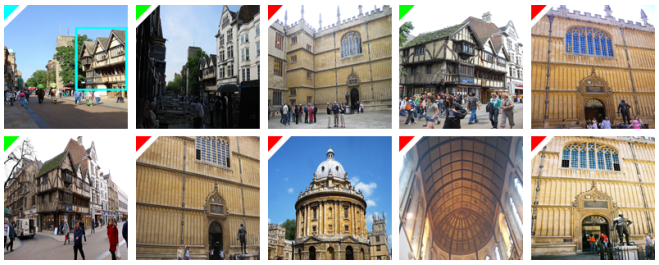


**Fig. 1**. A query from Oxford5K and its positive and negative pivots obtained with ASMK$^\star$. Some negative images are from the same object class, validating the assumption in Section 2.2.

QE is realized in this way such that a database image is determined as negative unless it is similar to the majority of pivots. The problem now is how to appropriately define the distribution $p(\mathbf{z})$ in Eq. 2.

### 2.2. Diffusion

We assume that 1) the pivots, which are on the same manifold composing the majority of the pivot set, are more likely to be positive and 2) visually similar pivots are more likely to have similar probabilities of being positive. We can thus spread the query similarities over the manifolds captured by the pivot set to learn the distribution $p(\mathbf{z})$. The above assumptions are usually reasonable since we have constrained the diffusion semi-locally with the pivots. In this paper, we mainly follow Zhou et al. [13] as regards diffusion.

We define an affinity matrix $\mathbf{A} = \{a_{ij}\}$ as $a_{ij} = s(\mathbf{z}_i, \mathbf{z}_j)$ for all $\mathbf{z}_i, \mathbf{z}_j \in \mathcal{Z}$. $\mathbf{A}$ is constrained to have zero self-similarities such that $\mathrm{diag}(\mathbf{A}) = \mathbf{0}$. It can be understood as the affinity matrix of a weighted graph $\mathcal{G}$ with vertices $\mathcal{Z}$. We then draw a degree matrix $\mathbf{D} = \mathrm{diag}((d_0, d_1, \ldots, d_m)^\top)$ from $\mathbf{A}$, in which $d_j$ denotes each column-wise sum of $\mathbf{A}$. Let $\mathbf{y} = \{y_j\} \in \mathbb{R}^{m+1}$ be an $l^1$-unit vector of query similarities with $y_j \propto s(\mathbf{q}, \mathbf{z}_j)$. The objective is to obtain a weight $w_j$ for each $\mathbf{z}_j$, represented by a vector $\mathbf{w} = \{w_j\} \in \mathbb{R}^{m+1}$. This is formulated by minimizing the cost function in Eq. 3.

$$\begin{aligned}\mathbf{w}^\star &= \arg\min_{\mathbf{w}} \left( \frac{1}{2} \sum_{i,j=0}^{m} a_{ij} \left( \frac{w_i}{d_i} - \frac{w_j}{d_j} \right)^2 + \mu \sum_{j=0}^{m} \frac{(w_j - y_j)^2}{d_j} \right) \\ &= \beta(\mathbf{I} - \alpha \mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{y} \end{aligned} \tag{3}$$

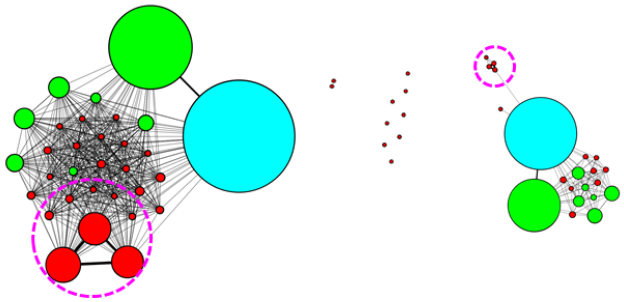
Here, $\mu$ is a positive regularization parameter, $\alpha = \frac{1}{1+\mu}$, $\beta = 1 - \alpha$, and $\mathbf{I}$ is a unit matrix. The first term in the cost function formulates the constraints of the assumptions described above; the second term imposes a fitting constraint on the solution such that the optimized $\mathbf{w}^\star$ does not vary greatly from $\mathbf{y}$.

It is usual to approximate Eq. 3 with a power method [13] for greater efficiency. Let $\mathbf{P} = \mathbf{A}\mathbf{D}^{-1}$. Let $\mathbf{w}^t$ be the vector updated in the $t$-th iteration and initialize $\mathbf{w}^0 = \mathbf{y}$. We have

$$\mathbf{w}^t = \alpha \mathbf{P} \mathbf{w}^{t-1} + (1 - \alpha)\mathbf{y}. \tag{4}$$

We directly define $p(\mathbf{z}_j) = w_j^\star$ and substitute it into Eq. 2 to give the final ranking list of $\mathbf{q}$. Since $\mathbf{P}$ is a column-wise normalized matrix and $\mathbf{y}$ is an $l^1$-unit vector, $\mathbf{w}^\star$ optimized by either Eq. 3 or Eq. 4 has a ready interpretation as probability, satisfying $\|\mathbf{w}^\star\|_1 = 1$.

In practice, the assumptions described above are reasonable in almost but not quite all cases. Some negative images may occupy the majority of pivots as shown in Fig. 1, and thus mislead the diffusion into an irrelevant manifold. To handle such outliers, we incorporate a local constraint in Section 2.3, and propose a mutual rank graph in Section 2.4 to account for varying local density in the image space.



(a) Complete graph (b) Mutual $k$-NN graph ($k = 30$)

**Fig. 2**. A query from Oxford5K and its positive and negative pivots connected in (a) a complete graph and (b) a mutual $k$-NN graph. Edge thickness indicates ASMK$^\star$ similarity and vertex size $p(\mathbf{z})$ learned with diffusion. The graph is generated using Gephi [22].

### 2.3. Local Constraint

We impose a local constraint [18] on the diffusion: only pairs of pivots that are reciprocal (mutual) nearest neighbors are kept as edges in the graph. Given a pivot $\mathbf{z}$, let its $k$-nearest neighbors ($k$-NNs) among $\mathcal{X}$ be $\mathcal{N}_k(\mathbf{z})$. We then modify $s(\cdot, \cdot)$ such that

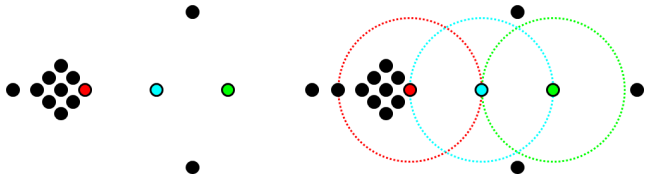
$$s_k(\mathbf{z}_i, \mathbf{z}_j) = \begin{cases} s(\mathbf{z}_i, \mathbf{z}_j) & \text{if } \mathbf{z}_i \in \mathcal{N}_k(\mathbf{z}_j) \text{ and } \mathbf{z}_j \in \mathcal{N}_k(\mathbf{z}_i) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This modified similarity $s_k(\cdot, \cdot)$ defines a mutual $k$-NN graph. In this way, the risk that an irrelevant class coexists in the same manifold with the query can be mitigated because different classes are usually distributed in the image space with different scales. Figure 2 shows an example where the mutual $k$-NN graph successfully separates a large proportion of false positives from the query manifold. The pivots surrounded by the pink circle establish a tightly settled submanifold in Fig. 2a and are redundantly connected with other pivots. Their weights are also greatly depressed in Fig. 2b by employing the local constraint.

Since the probability $p(\mathbf{x})$ for a database image is obtained by marginalizing over $p(\mathbf{z})$ instead of diffusion, imposing the local constraint on $p(\mathbf{x}|\mathbf{z})$ may lead to over-constrained QE. Hence, we keep Eq. 1 unchanged and use $s_k(\cdot, \cdot)$ only to construct $\mathbf{A}$ and $\mathbf{y}$ such that $a_{ij} = s_k(\mathbf{z}_i, \mathbf{z}_j)$ and $y_j \propto s_k(\mathbf{q}, \mathbf{z}_j)$.

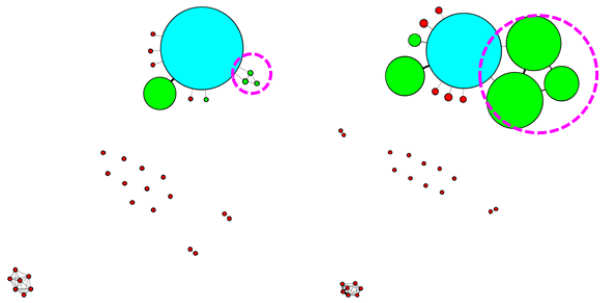
### 2.4. Mutual Rank Graph

We have yet to mention the definition of $s(\cdot, \cdot)$. In the diffusion process, a Gaussian kernel parameterized by a scaling factor $\sigma$ is the most widely used. When the data include clusters with different local statistics, it becomes difficult to select a single $\sigma$ that works well for all the data. A toy example is shown in Fig. 3a where the green and red circles are treated equally in terms of query affinity. Zelnik-Manor and Perona [14] proposed defining a local scaling (LS) factor for each datum in terms of its distance to its $k_\sigma$-th NN. When incorporated in a mutual $k$-NN graph, the method shows significant dependency on the neighborhood size $k$, resulting in unstable performance as $k$ increases.



(a) Affinity as Gaussian kernel (b) Affinity based on mutual rank

**Fig. 3**. Points in descriptor space. Tied NNs of query are in red and green. Dashed circles show their smallest discs containing the query. (a) Affinity as Gaussian kernel treats the two points equally. (b) We weight the green point more because of its higher mutual rank.



(a) Affinity as Gaussian kernel (b) Mutual rank graph

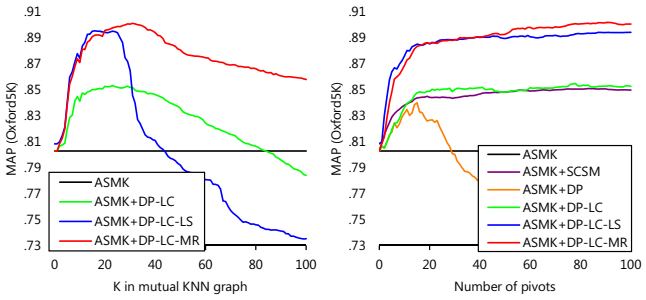
**Fig. 4**. A query from Oxford5K and its positive and negative pivots. (a) ASMK$^\star$ similarity and (b) the reciprocal of mutual rank are indicated with edge thickness. Vertex size indicates $p(\mathbf{z})$ learned with diffusion. $k = 30$. The graph is generated using Gephi [22].

Instead of selecting a Gaussian kernel, we propose defining the affinity between two images on the basis of their mutual rank. Let $r(\mathbf{z}_j|\mathbf{z}_i)$ be the rank of $\mathbf{z}_j$ when the database $\mathcal{X}$ is queried by $\mathbf{z}_i$. We can calculate an affinity $s_b(\cdot,\cdot)$, indicating the bidirectional relationship implied by mutuality:

$$s_b(\mathbf{z}_i, \mathbf{z}_j) = \left( \frac{[r(\mathbf{z}_j|\mathbf{z}_i)]^b + [r(\mathbf{z}_i|\mathbf{z}_j)]^b}{2} \right)^{-1/b}. \quad (6)$$

This measure is actually the reciprocal of the generalized mean between $r(\mathbf{z}_j|\mathbf{z}_i)$ and $r(\mathbf{z}_i|\mathbf{z}_j)$. It gives an advantage to the affinity of points with low local descriptor densities. As shown in Fig. 3b, our method weights the green point more because it has a higher mutual rank with respect to the query. The query affinity of the red point is decreased because it resides in a more densely populated region and is less tightly related to the query. Similar ideas have been used for retrieval problems [10, 11], but not for diffusion to the best of our knowledge. The reciprocal of Eq. 6 moves away from the geometric mean toward the lower rank, i.e., the larger $r$ value, for an increasing $b \in [0, \infty)$. We focus on $s_b(\cdot,\cdot)$ with $b \to \infty$ hereafter because the lower rank is more discriminating in suppressing outliers.

We incorporate Eq. 6 in Eq. 5, replacing $s(\cdot,\cdot)$ with $s_b(\cdot,\cdot)$, to construct $\mathbf{A}$ and $\mathbf{y}$. A mutual $k$-NN graph is thus adapted to a mutual rank graph, on which the distribution $p(\mathbf{z})$ is learned with diffusion. Figure 4 compares the two graphs. The pivots surrounded by the pink circle are underrated in Fig. 4a because of their smaller query similarities. However, the mutual rank graph in Fig. 4b successfully captures their low local image-space densities and greatly increases




(a) (b)

**Fig. 5**. Impact of $k$ used in mutual $k$-NN and $m$ as number of pivots. (a) $k_\sigma = 190$ for LS and $m = 100$ for all methods. (b) $k = 20$ for LC, $k = 20$ and $k_\sigma = 190$ for LS, and $k = 30$ for MR.

their weights such that these positive pivots can contribute more to QE. The reciprocal of mutual rank $s_b(\mathbf{x}, \mathbf{z})$ is also employed in Eq. 1 to provide higher discriminative power.

## 3. EXPERIMENTS

### 3.1. Experimental Setup

We used three benchmarks that are widely used in image retrieval: Oxford5K [23], Oxford105K [23], and Paris6K [24]. Mean average precision (MAP) was used as a performance measure for all datasets.

An aggregated selective match kernel (ASMK$^\star$) [25] was employed as the primary image similarity. The visual vocabulary of Oxford5K was trained on Paris6K and vice versa. We did not investigate the compatibility of our method with other metrics, but any state-of-the-art image representations, e.g., CNN features [28, 29], can be used here. The damping factor $\alpha$ is always 0.99 as in the work of Zhou et al. [13]. 30 iterations were found to be sufficient for the power method to yield a good solution.

During the offline stage, ASMK$^\star$ similarities $s(\cdot,\cdot)$ and unidirectional ranks $r(\cdot|\cdot)$ between all $\mathbf{x} \in \mathcal{X}$ were computed in advance. The similarities were represented by a set of $n$ sorted lists, each having a length of $n$. We thus have two $n \times n$ intermediate matrices. On a large scale, e.g., for Oxford105K, we truncated the matrices, only keeping the 5,000 NNs for each $\mathbf{x}$ to realize a lower space requirement. To conduct an online search, the similarities between the query $\mathbf{q}$ and all $\mathbf{x} \in \mathcal{X}$ were computed and $m$ pivots were selected. The affinity matrix $\mathbf{A}$ and the vector $\mathbf{y}$ were constructed with Eqs. 5 and 6 for diffusion. The conditional probabilities $p(\mathbf{x}|\mathbf{z})$ could be easily obtained from the rank matrix except for $\mathbf{z}_0$: we approximated $s_b(\mathbf{q}, \mathbf{x})$ by assuming $r(\mathbf{q}|\mathbf{x}) = 1$ for lower online cost.

### 3.2. Parameter Investigation

We varied $k$ used in the mutual $k$-NN graph and the number of pivots $m$, and evaluated the performance for different configurations. Results for Oxford5K are shown in Fig. 5. Here, DP realizes QE with a diffusion process on a complete graph where the edges are weighted by ASMK$^\star$ similarities; LC [18] replaces the complete graph with a mutual $k$-NN graph; LS weights each edge with a locally scaled Gaussian kernel [14], in which the Euclidean distance is substituted by the negative log-likelihood of the ASMK$^\star$ similarity; MR weights each edge with the reciprocal of mutual rank, which corresponds to our method.

**Table 1**. Comparison with studies based on local features. Results based on our reimplementation are marked with †, otherwise they are from original papers. SP indicates spatial verification. The previous highest and our highest MAPs are underlined and shown in bold, respectively.

| | QE | SP | $m$ | MAP (%) | | |
|---|---|---|---|---|---|---|
| | | | | Oxford5K [23] | Oxford105K [23] | Paris6K [24] |
| ASMK⋆ [25]† | ✗ | ✗ | – | 80.3 | 74.9 | 77.1 |
| HQE [12] | ✓ | ✗ | 100 | 83.8 | 80.4 | 82.8 |
| Mikulík et al. [26] | ✓ | ✓ | ∼ 50 | 84.9 | 79.5 | 82.4 |
| Qin et al. [27] | ✓ | ✗ | – | 85.0 | 81.6 | 85.5 |
| ASMK⋆+SCSM [11]† | ✓ | ✗ | 100 | 85.0 | 82.0 | 84.4 |
| i-ASMK⋆+HQE [25] | ✓ | ✗ | 100 | 86.9 | 85.3 | 85.1 |
| HQE-SP [12] | ✓ | ✓ | 100 | 88.0 | 84.0 | 82.8 |
| ASMK⋆+DP-LC-MR | ✓ | ✗ | 25 | 88.8 | 87.1 | 85.6 |
| ASMK⋆+DP-LC-MR | ✓ | ✗ | 50 | 89.3 | 88.0 | 87.3 |
| ASMK⋆+DP-LC-MR | ✓ | ✗ | 100 | **90.1** | **88.6** | **89.0** |


**Fig. 6**. Query examples from Oxford5K and Paris6K and difficult instances low-ranked by ASMK⋆, i.e., the baseline. Precision at the position where each instance is retrieved is shown under each instance for ASMK⋆ without and **with** our QE method.

As shown in Fig. 5a, the performance stays stable over a wide $k$ range for LC and our method. In comparison, LS is very sensitive in terms of this parameter. Considering the additional parameter $k_\sigma$, which also has to be tuned, LS is difficult to generalize for different data. In Fig. 5b, DP suffers from significant query drift as $m$ increases, justifying our motivation for the incorporation of the stronger local constraints described in Sections 2.3 and 2.4. LC by itself already solves a major part of this problem. Combining it with the mutual rank graph provides yet another substantial improvement in MAP of up to 5%. For Oxford5K, the precision of ASMK⋆ at the position where the 100th pivot is selected is 32.7% on average with a minimum value of 6%. Even with so few positives in the pivot set, our method still achieves a long-lasting improvement, demonstrating its great robustness as regards query drift. We did not tune the parameters on Oxford105K or Paris6K but simply set $k = 30$ for the rest of this paper.

### 3.3. Comparison

We compared our method with the state-of-the-art techniques based on handcrafted features. Table 1 summarizes the results, where the previous highest MAPs are underlined. With only 25 pivots, our method already outperforms all other methods in all datasets. Moreover, our method does not rely on spatial verification, which usually involves additional online costs. Figure 6 shows qualitative results revealing that images benefit from our method mainly when they vary from the depicted object with large illumination and viewpoint change. For Oxford105K, our method took 19.2 hours for database side ASMK⋆ computation and 48 minutes for sorting and truncation. Both of these processes were finished offline and can be performed in parallel for increased speed. The query time with the same dataset is 1.33s on average for a primary search based on ASMK⋆ and 0.76s for QE with 100 pivots. These experimental results demonstrate the great scalability of our method.

## 4. CONCLUSION

We proposed a QE method that handles the danger of query drift by investigating manifolds of pseudo-relevant images in a topological space. The method yields greater robustness than previous studies especially when false positives reside on different manifolds from the query. We show experimentally that it is less parameter-sensitive and consistently outperforms state-of-the-art techniques based on handcrafted features. Our method is scalable since no large-scale matrix manipulations or iterations are required online. Unlike previous studies [5–7, 12] that realize QE at the feature level, our method does not depend on specific image representations. As a future task, we shall verify this experimentally and investigate the compatibility of our method with recent CNN descriptors [28, 29].

## 5. REFERENCES

[1] Tobias Weyand and Bastian Leibe, "Discovering favorite views of popular places with iconoid shift," in *ICCV*, 2011, pp. 1132–1139. 1

[2] Relja Arandjelovic, Petr Gronát, Akihiko Torii, Tomás Pajdla, and Josef Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *CVPR*, 2016, pp. 5297–5307. 1

[3] Jared Heinly, Johannes L. Schönberger, Enrique Dunn, and Jan-Michael Frahm, "Reconstructing the world* in six days," in *CVPR*, 2015, pp. 3287–3295. 1

[4] Johannes L. Schönberger, Filip Radenovic, Ondrej Chum, and Jan-Michael Frahm, "From single image query to detailed 3D reconstruction," in *CVPR*, 2015, pp. 5126–5134. 1

[5] Ondrej Chum, James Philbin, Josef Sivic, Michael Isard, and Andrew Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in *ICCV*, 2007, pp. 1–8. 1, 4

[6] Herve Jegou, Hedi Harzallah, and Cordelia Schmid, "A contextual dissimilarity measure for accurate and efficient image search," in *CVPR*, 2007. 1, 4

[7] Ondrej Chum, Andrej Mikulík, Michal Perdoch, and Jiri Matas, "Total recall II: Query expansion revisited," in *CVPR*, 2011, pp. 889–896. 1, 4

[8] Danfeng Qin, Stephan Gammeter, Lukas Bossard, Till Quack, and Luc J. Van Gool, "Hello neighbor: Accurate object retrieval with $k$-reciprocal nearest neighbors," in *CVPR*, 2011, pp. 777–784. 1

[9] Relja Arandjelovic and Andrew Zisserman, "Three things everyone should know to improve object retrieval," in *CVPR*, 2012, pp. 2911–2918. 1

[10] Agni Delvinioti, Hervé Jégou, Laurent Amsaleg, and Michael E. Houle, "Image retrieval with reciprocal and shared nearest neighbors," in *VISAPP*, 2014, pp. 321–328. 1, 3

[11] Xiaohui Shen, Zhe Lin, Jonathan Brandt, and Ying Wu, "Spatially-constrained similarity measure for large-scale object retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1229–1241, 2014. 1, 3, 4

[12] Giorgos Tolias and Hervé Jégou, "Visual query expansion with or without geometry: Refining local descriptors by feature aggregation," *Pattern Recognition*, vol. 47, no. 10, pp. 3466–3476, 2014. 1, 4

[13] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf, "Learning with local and global consistency," in *NIPS*, 2003, pp. 321–328. 1, 2, 3

[14] Lihi Zelnik-Manor and Pietro Perona, "Self-tuning spectral clustering," in *NIPS*, 2004, pp. 1601–1608. 1, 2, 3

[15] Leo Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006. 1

[16] Song Lu, Vijay Mahadevan, and Nuno Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *CVPR*, 2014, pp. 2790–2797. 1

[17] Sravan Gudivada and Adrian G. Bors, "Ortho-diffusion decompositions of graph-based representation of images," *Pattern Recognition*, vol. 48, no. 12, pp. 4097–4115, 2015. 1

[18] Peter Kontschieder, Michael Donoser, and Horst Bischof, "Beyond pairwise shape similarity analysis," in *ACCV*, 2009, pp. 655–666. 1, 2, 3

[19] Xingwei Yang, Suzan Köknar-Tezel, and Longin Jan Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," in *CVPR*, 2009, pp. 357–364. 1

[20] Michael Donoser and Horst Bischof, "Diffusion processes for retrieval revisited," in *CVPR*, 2013, pp. 1320–1327. 1

[21] Ahmet Iscen, Giorgos Tolias, Yannis S. Avrithis, Teddy Furon, and Ondrej Chum, "Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations," in *CVPR*, 2017, pp. 926–935. 1

[22] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *ICWSM*, 2009. 2, 3

[23] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *CVPR*, 2007. 3, 4

[24] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *CVPR*, 2008. 3, 4

[25] Giorgos Tolias, Yannis S. Avrithis, and Hervé Jégou, "Image search with selective match kernels: Aggregation across single and multiple images," *International Journal of Computer Vision*, vol. 116, no. 3, pp. 247–261, 2016. 3, 4

[26] Andrej Mikulík, Michal Perdoch, Ondrej Chum, and Jiri Matas, "Learning vocabularies over a fine quantization," *International Journal of Computer Vision*, vol. 103, no. 1, pp. 163–175, 2013. 4

[27] Danfeng Qin, Christian Wengert, and Luc J. Van Gool, "Query adaptive similarity for large scale object retrieval," in *CVPR*, 2013, pp. 1610–1617. 4

[28] Filip Radenovic, Giorgos Tolias, and Ondrej Chum, "CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples," in *ECCV*, 2016, pp. 3–20. 3, 4

[29] Albert Gordo, Jon Almazán, Jérôme Revaud, and Diane Larlus, "End-to-end learning of deep visual representations for image retrieval," *International Journal of Computer Vision*, vol. 124, no. 2, pp. 237–254, 2017. 3, 4