Pairwise Geometric Matching for Large-scale Object Retrieval

Xinchao Li, Martha Larson, Alan Hanjalic
Multimedia Computing Group, Delft University of Technology
Delft, The Netherlands
{x.li-3,m.a.larson,a.hanjalic}@tudelft.nl

Abstract

Spatial verification is a key step in boosting the performance of object-based image retrieval. It serves to eliminate unreliable correspondences between salient points in a given pair of images, and is typically performed by analyzing the consistency of spatial transformations between the image regions involved in individual correspondences. In this paper, we consider the pairwise geometric relations between correspondences and propose a strategy to incorporate these relations at significantly reduced computational cost, which makes it suitable for large-scale object retrieval. In addition, we combine the information on geometric relations from both the individual correspondences and pairs of correspondences to further improve the verification accuracy. Experimental results on three reference datasets show that the proposed approach results in a substantial performance improvement compared to the existing methods, without making concessions regarding computational efficiency.

1. Introduction

In this paper we address the challenge of improving the efficiency and reliability of image matching in an object-based image retrieval scenario. Under object-based image retrieval, further referred simply to as "object retrieval", we understand the problem of finding images that contain the same object(s) or scene elements as in the query image, however, possibly captured under different conditions in terms of rotation, viewpoint, zoom level, occlusion or blur. Many object retrieval approaches and methods [9, 22, 13, 1, 27] have been proposed in recent literature, largely inspired by the pioneering work of Sivic and Zisserman [24] and built on the bag-of-features (BOF) principle for image representation. An analysis of the state-of-the-art reveals that these approaches and methods are typically centered around the idea of detecting and verifying correspondences between salient points in a given pair of images. The initial set of correspondences are detected based on matches between visual feature statistics measured in different images around found salient points. The correspondence verification step then serves to filter out unreliable correspondences. This verification is typically a spatial (geometric) one and involves geometric constraints to secure consistency of transformation of different image points. Spatial verification is the key to achieve high precision for object retrieval, especially when searching in large, heterogeneous image collections [24, 21].

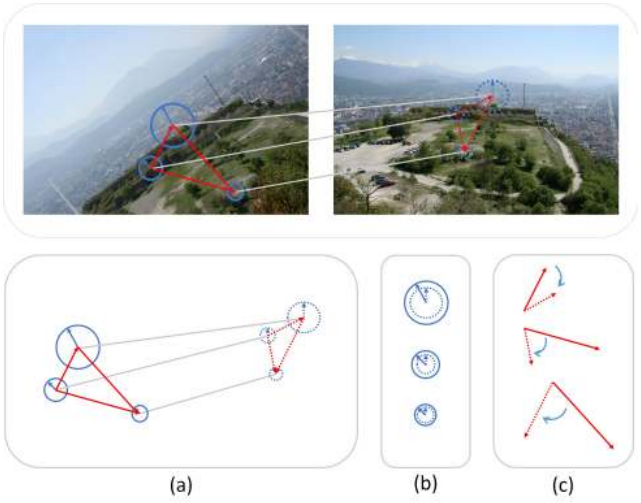


Figure 1: (a) Three correspondences found for two images, (b) global rotation and scale relations between images encoded in the transformation of the matched salient points from individual correspondences, (c) rotation and scale relations between vectors formed by pairwise salient points involved in the correspondences. Transformations in cases (b) and (c) are closely related to each other and can be used to emphasize each other for spatial verification.

A common way of verifying the initial correspondences is to apply a *geometric matching*. Geometric matching can be done either explicitly, by iteratively building an optimized transformation model and fitting it to the initial correspondences (*e.g.*, RANSAC-based model fitting ap-

proaches [21, 7]), or implicitly, *e.g*., by verifying the consistency of the image points involved in the correspondences in the Hough transform space [14, 2]. Compared to these approaches, pairwise relative geometric relations between the correspondences have not been frequently exploited for spatial verification. This may be due to the fact that the typical number N of initially detected correspondences is usually large, resulting in high computational complexity of pairwise comparisons, which can be modeled as $\mathcal{O}(N^2)$. This complexity makes exploitation of pairwise relations less attractive when operating on large image collections. Exploiting these pairwise geometric relations could, however, further improve the performance of image matching as it brings valuable additional information about local object or scene constraints of the correspondences into the matching process. As illustrated in Figure 1, the geometric relations in terms of rotation and scaling between vectors formed by a pair of correspondences are closely related to the global geometric relations between images that are encoded in the transformation of the image regions surrounding the salient points. Our goal in this paper is therefore twofold. First, we aim at generating the conditions under which pairwise geometric relations can be applied for spatial verification of correspondences at a reasonable computational cost. Second, we aim at maximizing the benefit of involving these relations for improving the object retrieval performance.

We pursue the goal specified above by a novel *pairwise geometric matching* method that consists of three main steps. We first propose a one-versus-one ('1vs1') matching strategy for the initial correspondence set to handle the redundancy of one-to-many correspondences, which is a typical result of detecting correspondences between two images [13] [2]. By removing this redundancy, a new, significantly reduced correspondence set is generated. Then, similarly to [17, 14], we reduce this set even further, by deploying Hough voting in the scaling and rotation transformation space. After these two steps, a large fraction of original correspondences are filtered out, which enables us to exploit pairwise geometric relations for spatial verification at a significantly reduced computational cost. Finally, a simple pairwise weighting method is devised to incorporate both the global geometric relations derived from individual correspondences and the local pairwise relations of pairs of correspondences. As we will show by experimental results in Section 6, our proposed method makes the spatial verification of correspondences more tractable in case of a large image collection, but also more reliable, which leads to an overall significant improvement of the object retrieval performance compared to state-of-the-art methods.

2. Related Work and Contribution

The existing work addressing the problem of verifying the geometric consistency within a set of correspondences can be grouped in two main categories. The first category comprises the methods exploiting *individual* point correspondences for spatial verification, while the methods from the second category exploit multiple correspondences for this purpose. We briefly analyze the representative methods from these categories and position our contribution with respect to them.

2.1. Exploiting individual correspondences

Model-based methods. For two images capturing the same object, a limited number of correspondences can be deployed to estimate the geometric model transforming the points of one image into those of the other image [11]. Once the model is obtained, each correspondence can be assessed in how it fits this model. The key challenge here is how to do model estimation in the presence of noisy correspondences. One of the classical methods to pursue this challenge is RANSAC [10]. Over the years, several attempts have been made to improve its efficiency. For example, Chum *et al*. [7] managed to significantly speed up the model estimation by adding a generalized model optimization step when the new maximum of inliers is reached. This results in less iterations needed for model estimation to converge. Philbin *et al*. [21] exploited local appearance of matched image points to generate model hypotheses using a single correspondence, which significantly reduces the amount of possible model hypotheses. Different from RANSAC-based methods, Lowe [17] applied Hough transform to the geometric transformation space to find groups of consistently transformed correspondences prior to estimating the transformation model. In contrast to these model-based methods, which typically need complex iterative model optimization, we are targeting a more lightweight, model-free method.

Model-free methods. As an alternative to the methods discussed above, one can also implicitly verify the correspondences with respect to their consistency in the Hough transformation space. Avrithis and Tolias [2] exploited the relative geometric relations, *i.e*., scaling, orientation and location, between the local appearance of the matched points. Each correspondence generates one vote in the 4-dimensional transformation space and is then weighted by pyramid matching to capture its consistency with other correspondences. Jégou *et al*. [14] used the scaling and orientation relations between matched points to find the correspondences that agree with the dominant transformation found in the transformation space. Similarly, Zhang *et al*. [28] exploited the translation between matched points using Hough voting in a 2-dimensional translation space. Shen *et al*. [23] also exploited the translation using Hough voting. However, instead of using only the original query object, they applied several transformations with different rotations and scales to the query object, and searched for the best possible translation of these transformed query objects against a collection

image. In this way, rotation and scaling invariance can be added to the system. Our proposed method belongs to this category of model-free approaches. However, in contrast to most of the existing work in this direction, which focuses on individual correspondences, we are considering the pairwise relations between correspondences as well.

2.2. Exploiting multiple correspondences

In contrast to rich previous work focusing on individual correspondences, the information encoded in groups of correspondences has remained less exploited for spatial verification. Some related methods implicitly encode the spatial-order information of the correspondences. Wu *et al*. [26] bundled the local features according to their location and captured the relative order consistency of the correspondences along the X- and Y-coordinates in each image. As this simple way of capturing order consistency cannot support complex geometric transformations, it is primarily suitable for problems of near-duplicate detection. Compared to this, Cao *et al*. [4] encoded the spatial-order relation between local features by ordering them in a set of linear and circular directions, so rotation can be handled as well. Instead of relying on the ordering of the correspondences, we deploy a more subtle information for spatial verification, namely the rotation and scaling relations between the vectors formed by salient points involved in correspondences. This is likely to make spatial verification more reliable.

We are not the first ones exploiting pairwise geometric relations between correspondences. Carneiro and Jepson [6] employed a pairwise semi-local spatial similarity to capture the pairwise relations of correspondences and grouped them using connected component analysis based on the pairwise similarity matrix. This work was further combined with a probabilistic verification method in [5] to increase the proportion of correct matches in the correspondence set. Likewise, by building a pairwise similarity matrix of correspondences, Leordeanu and Hebert [16] employed a spectral method to greedily recover inliers and find the strongly connected cluster within the correspondence set. These works are related to our approach as they all exploit the pairwise relation between correspondences. However, these methods were designed to exploit the pairwise relations directly from the initial correspondences. As discussed earlier in this paper, the complexity of spatial verification in this case becomes too high to be applicable in the case of a large image collection. Compared to these methods, our contribution is twofold. First, we significantly reduce the number of correspondences and in this way make the proposed spatial verification more tractable. Second, our pairwise geometric matching method combines both the global geometric relations derived from individual correspondences and the local pairwise relations of pairs of correspondences for improved object retrieval performance.

3. Correspondence problem formulation

We start out from a standard representation of an image using local features. This representation typically involves detection of salient points in the image and representation of these points by suitable feature vectors describing local image regions around these points. For instance, in the SIFT [17] scheme, which is widely deployed for this purpose, salient points are detected by a Difference of Gaussians (DOG) function applied in the scale space. The points are then represented by local feature vectors $\mathbf{f} = [\mathbf{x}, \theta, \sigma, \mathbf{q}]$, where $\mathbf{x}$, θ and σ stand for the spatial location, dominant orientation and scale of the represented region around the point, respectively, and $\mathbf{q}$ is the feature description of the region. Given the images F and $\tilde{F}$, and their salient points with indexes i and m and represented by feature vectors $\mathbf{f}_i$ and $\tilde{\mathbf{f}}_m$, respectively, we define the initial set $\mathbf{C}$ of correspondences c_{im} between them as

$$\mathbf{C} = \{(\mathbf{f}_i, \tilde{\mathbf{f}}_m, W_{ini}(c_{im}) | \Phi(\mathbf{f}_i, \tilde{\mathbf{f}}_m) = 1\} \qquad (1)$$

Here, $\Phi(.) \in \{0, 1\}$ is the binary matching function serving to judge whether two image points capture the same object point in the physical world. For instance, in the BOF scheme, this function is typically computed as $\Phi = \delta(u(\mathbf{q}_i) - u(\tilde{\mathbf{q}}_m))$, where $u(\mathbf{q}_i)$ is the quantized cluster center of the description vector $\mathbf{q}_i$ of local feature $\mathbf{f}_i$ and where $\delta(.)$ is the Kronecker delta. Furthermore, $W_{ini}(c_{im})$ is the weight initially assigned to a correspondence c_{im} and representing the proximity between two points in the local feature space. For instance, the weight can be computed in terms of the statistical distinctiveness of the quantized visual feature center within the image collection, *e.g*., using the inverse document frequency (*idf*) scheme applied in the BOF context [24]. As an alternative, this weight can also be computed using Hamming distance employed in the Hamming Embedding scheme [13, 14].

4. Pairwise Geometric Matching

In this section we describe the three steps of our proposed pairwise geometric matching method: (a) applying the '1vs1' matching constraint, (b) Hough voting and (c) integrating global and pairwise geometric relations.

4.1. 1vs1 matching

The initial correspondence set $\mathbf{C}$ usually contains a large portion of outliers, or incorrect correspondences, and can include multiple mappings for one single point, *i.e*., the burstiness phenomenon observed in [13]. However, object matching implies that one object point in one image can only have one corresponding point in another image. Therefore, the final verified correspondence set should only contain unique correspondences between points.

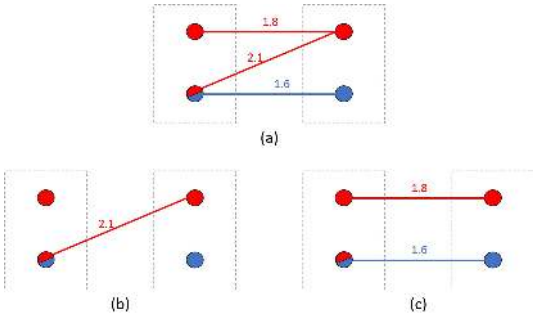


Figure 2: Illustration of two different strategies for filtering out multiple alternative correspondences. Case (a) shows the original correspondences. The lower point in the query image (left image) represents a point that matches two different points marked with red and blue in the right image. Case (b) illustrates the strategy by Jegou et al. [13] that focuses on the strongest correspondences. Case (c) is the proposed '1vs1' strategy that balances filtering out of the correspondences with preserving as many informative correspondences as possible.

To achieve this, one can formulate an assignment problem, where one can minimize the overall distance between two point sets by using the Hungarian algorithm with the computing time in $O(N^3)$ for set with N features [15]. As finding optimal matches is time consuming, one can aim at an approximate solution. For instance, Jégou et al. [13] proposed to choose the strongest match per point first and then discard all the other matches associated with matched points. However, as can be seen from Figure 2 (case (b)), this strategy may result in insufficient number of matches for geometric check. In order to generate a more robust solution, we devise the '1vs1' matching strategy and apply it to the initial correspondence set $\mathbf{C}$.

As illustrated by the case (c) in Figure 2, in our approach we focus on preserving as many correspondences as possible to maximally inform the assessment of the relation between two images. We first start from the point that originally has fewest matching correspondences assigned (i.e., potential unique matches), select the one with the highest weight, and then discard other matches that contain points belonging to this selected correspondence. We continue this process until no more points need to be processed. In this way, we generate a correspondence set $\mathbf{C}_{1vs1}$ that serves as input for further steps.

4.2. Hough voting

We now depart from the set $\mathbf{C}_{1vs1}$ and follow the strategy from [17, 14] to apply a Hough voting scheme in search for dominant ranges of the target transformation parameters, specifically for the rotation and scaling, in the transformation space. Then, we further reduce the number of correspondences by filtering out those that are not consistently transformed within these ranges.

Each correspondence, c_{im}, stands for a transformation from point i in image F to point m in image $\tilde{F}$. The rotation and scaling relations for this correspondence are denoted, respectively, by

$$\theta = \theta_m - \theta_i, \;\; \sigma = \sigma_m / \sigma_i \tag{2}$$

Each correspondence gives a vote in the 2-dimensional rotation-scaling transformation. The dominant ranges of these two transformation parameters, denoted as B_ϑ and B_ς, emerge as the corresponding ranges of the largest bin in the 2-dimensional voting histogram. The correspondences with votes falling in this largest bin are considered to most reliably reveal the transformation between two images. They form the set $\mathbf{C}_{R\&S}$, which serves as input into the last step of the proposed method.

4.3. Integrating global and pairwise geometric relations

We start out from the correspondences included in the set $\mathbf{C}_{R\&S}$ and assess the match between images F and $\tilde{F}$ based on pairwise geometric relations between the correspondences. These pairwise geometric relations are derived from the rotation and scaling relations between the corresponding vectors connecting the correspondences in the two images. Given the correspondences, c_g and c_h, which connect point i in image F to point m in image $\tilde{F}$, and point j in image F to point n in image $\tilde{F}$, respectively, we can generate vector $\mathbf{v}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ in image F and vector $\tilde{\mathbf{v}}_{mn} = \mathbf{x}_m - \mathbf{x}_n$ in image $\tilde{F}$. The pairwise geometric relations between the two vectors in terms of rotation and scaling can then be defined as

$$\begin{aligned} \theta_{gh} &= \arccos\left(\frac{\mathbf{v}_{ij} \cdot \tilde{\mathbf{v}}_{mn}}{||\mathbf{v}_{ij}|| \cdot ||\tilde{\mathbf{v}}_{mn}||}\right) \cdot \mathrm{sgn}(\mathbf{v}_{ij} \times \tilde{\mathbf{v}}_{mn}) \\ \sigma_{gh} &= \frac{||\tilde{\mathbf{v}}_{mn}||}{||\mathbf{v}_{ij}||} \end{aligned} \tag{3}$$

where θ_{gh} and σ_{gh} are the counterclockwise rotating angle and the scaling factor from $\mathbf{v}_{ij}$ to $\tilde{\mathbf{v}}_{mn}$, respectively.

Each correspondence c_g is then weighted by its pairwise rotation and scaling consistence with other correspondences:

$$W_{PG}(c_g) = \sum_{c_h \in \mathbf{C}_{R\&S}, h \neq g} f(\theta_{gh}, \sigma_{gh}) \tag{4}$$

where

$$f(\theta_{gh}, \sigma_{gh}) = \begin{cases} 1, & \text{if } \theta_{gh} \in B_\vartheta, \; \sigma_{gh} \in B_\varsigma \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

We note that the weights computed using Eq.4 combine together the information on geometric relations obtained

from individual correspondences, as imposed by the rotation and scale range limits B_ϑ and B_ς in Eq.5, and from the pairs of correspondences, as indicated by vector relations in Eq.3. The final matching score between two images is obtained as the sum of the weights $W_{PG}(c_g)$ of all correspondences from the set $\mathbf{C}_{R\&S}$:

$$S(F, \tilde{F}) = \sum_{c_g \in \mathbf{C}_{R\&S}} W_{PG}(c_g) \quad (6)$$

5. Experimental Setup

5.1. Object Retrieval Framework

We evaluate our proposed pairwise geometric matching method in an object retrieval context. For this purpose, we implemented an object retrieval system based on the classical bag-of-feature-based scheme [24] and considering recent advances in realizing this scheme [14, 13, 22]. To make the system scalable to large image collections, we implemented it using a Map-Reduce-based structure on a Hadoop-based distributed server[1].

Local descriptors and visual words: we use Hessian-affine detector [18] to detect salient points and compute SURF descriptors [3] for these points. As described in [2, 1], the bag-of-feature-based system performs differently depending on whether the visual words vocabulary is trained on an image set with or without test data, i.e., whether the vocabulary is *specific* or *generic*. To mimic the situation in a real retrieval system, we use a separate set of $50k$ randomly selected images from Flickr to learn the *generic* vocabulary set with exact k-means and use it in all experiments.

Weighting the initial correspondences and calculating initial ranking score: As indicated in Section 3, the initial set of correspondences can be weighted using different methods. We deploy two common weighting schemes:

(1) ***BOF***: We use the square of the inverse document frequency (*idf*) of the visual word associated with a correspondence as the matching weight. The initial ranking score for the retrieved images is obtained as the sum of the weights of all correspondences, divided by the L2 norm of the bag-of-feature vector.

(2) ***HE***: We employ the Hamming Embedding (HE)-based method proposed in [13] to weight the matched features based on the Hamming distance between their signatures. When calculating the initial ranking score, the burst weighting scheme developed in [13] is employed to handle the burstiness phenomenon in the initial ranking phase.

Multiple assignment (MA): To take into account the quantization noise introduced by a bag-of-feature image representation, we adopted the method from [13] to assign a descriptor to multiple visual words and applied it on the query side only to reduce the computational cost.

[1]This work was carried out on the Dutch national e-infrastructure with the support of SURF Foundation.

5.2. Experimental protocol

We assess the proposed method through a comparative experimental analysis and by following similar protocol and criteria as in [2]. We use the *precision-recall curve* to evaluate the pairwise image matching performance and use *mean average precision* (mAP) to evaluate the improvement in object retrieval using the proposed spatial verification method. In the experiments, we use three variants of our implemented object retrieval system based on the two weighting schemes introduced in Section 5.1: (1) ***BOF***, with a *generic* vocabulary of $100K$, as also deployed in [2], (2) ***HE***, with a *generic* vocabulary of $20K$ and with 64-bit Hamming signature and (3) ***HE+MA***, which is equivalent to *HE* combined with multiple assignment. This is the same setting as in [13]. We further denote our proposed pairwise geometric matching method as (***PGM***) and its three steps described in sections 4.1, 4.2 and 4.3 as ***1vs1***, ***HV*** and ***PG***, respectively. We refer to the three system realizations incorporating *PGM* as *BOF+PGM*, *HE+PGM* and *HE+MA+PGM*.

We compare these system realizations with state-of-the-art methods both integrally and by adding individual steps one by one in order to assess the contribution of each step to the overall object retrieval performance. We use three state-of-the-art methods as baselines that we refer to as ***HPM*** [2], ***SM*** [16] and ***FSM*** [21]. With respect to *HPM*, we do the comparison directly by integrating the binary code of [2] into our system. As this binary code does not support Hamming embedding, we only integrate it into the *BOF* setting, which is referred to as *BOF+HPM*. Regarding *SM* and *FSM*, as there were no original implementations available for them, the comparison is only indirect, using the experimental results reported in [2] that were obtained on the same datasets as in this paper.

5.3. Datasets

We conduct the experiments on three publicly available datasets commonly used in the related work, namely *Oxford* [21], *Holidays* [12] and *Barcelona* [25]. To mimic the large-scale image retrieval scenario, we follow the same strategy used in [13, 2] to add distractors to dataset images. We crawled 10 million geo-tagged photos from Flickr for this purpose. These distractors are distributed all around the world, except for Oxford and Barcelona regions.

6. Experiments

6.1. Impact of the parameters

We start our series of experiments by evaluating the impact of two main parameters, namely the bin sizes of rotation and scale used in Hough voting, on the system performance. These parameters control the trade-off between

filtering out the mismatches and remaining tolerant to non-rigid object deformations. We evaluate these parameters in the object retrieval scenario using the *HE+MA* system implementation. Based on the results in Table 1, we choose the bin size of 30 degrees for rotation and 0.2 for logarithmic scale as best performing across the two datasets and adopt these parameter values for all subsequent experiments.

Table 1: mAP comparison of *PGM* on *Oxford* and *Holidays* datasets with different bin sizes for rotation and scale.

	Oxford			Holidays		
	0.1	0.2	0.3	0.1	0.2	0.3
15	0.725	0.734	0.730	0.882	**0.893**	0.888
30	0.735	**0.737**	0.731	0.883	0.892	0.890
45	0.728	0.732	0.724	0.886	0.888	0.882

6.2. Pairwise image matching

To assess the *PGM* method, we follow the same experimental procedure as in [2], which enumerates all pairs of images in the *Barcelona* dataset and classifies each image pair to be relevant or irrelevant based on whether its matching score is higher than a threshold. There are in total 927 images in the *Barcelona* dataset, which form $927 \times 927 = 859329$ image pairs, and among which $74,075$ image pairs are relevant according to the ground truth. Figure 3 shows the precision-recall curves computed for various realizations of our system. Regarding the state-of-the-art, we compare our method directly with *HPM* and indirectly with *SM* based on the results reported in [2] and using similar basic system configuration. For recall of 0.9, *BOF+PGM* achieves the precision of 0.68, which is better than 0.42 achieved by *BOF+HPM* or 0.2 achieved by *SM*. We note that according to Figure 3, our method can achieve even better performance (precision of 0.83 at recall 0.9) if the best performing system variant is deployed.

6.3. Spatial verification for object retrieval

We now evaluate the proposed method in the object retrieval context. For each query image, top-1000 ranked images are selected to perform spatial verification. Since the rank order of these images is adjusted based on verification, we refer to this set of top-1000 images as the *reranking range*. We first evaluate *PGM* against the original datasets without distractors. According to Table 2, *PGM* clearly outperforms the baselines. Figure 4 shows examples of ranked images obtained using *PGM* and *HPM*.

Figure 5 illustrates the system performance with different sizes of image database. The binary code of *HPM* needs to keep all the index information in the memory, which in the case of a database of 10 million images, leads to memory consumption that is too large. For this reason, *HPM* is not included at this scale. The curves in the figure indicate the improvement of the performance after adding each of the steps of our method to the basic *BOF* system configuration. Step-for-step improvement is not clearly evident in the case of the *HE* system configuration. This is because in this configuration the 'burstiness' phenomenon is handled in the initial retrieval phase using burst weighting [13]. Therefore, the *1vs1* and *HV* steps cannot bring much additional improvement. *PG*, on the other hand, becomes the key step to improve over *HE*.

Regarding the comparison with the best performing baseline, *HPM*, we observe that *BOF+PGM* (cf. *+PG* in Figure 5) consistently outperforms *HPM* at each scale. Fur-

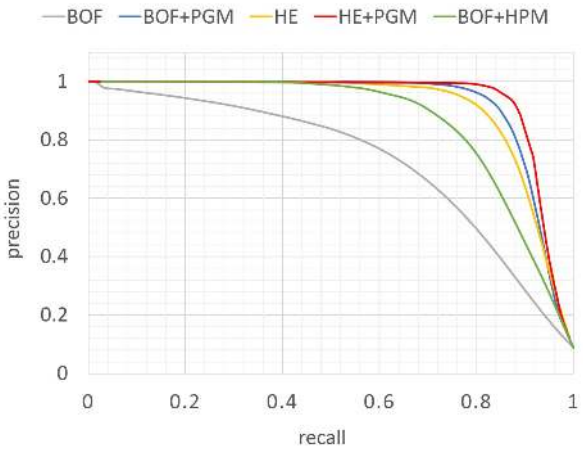


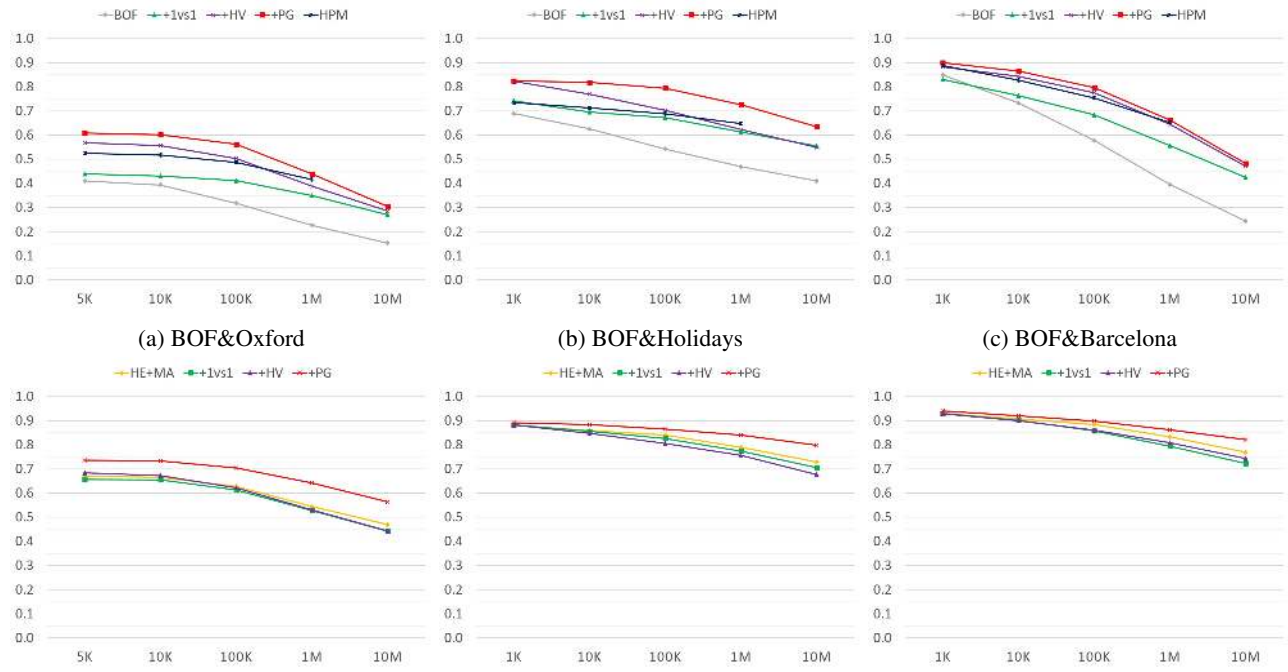
Figure 3: Precision-recall curves over all pairs of images in the *Barcelona* dataset.

Table 2: mAP comparison of different spatial verification schemes. All results are generated under the same conditions: reranking on top $1K$ ranked photos from BOF using SURF feature and *Single Assignment* on $100K$ vocabulary.

	FSM[1]	HPM[1]	HPM	PGM
Oxford	0.503	0.522	0.525	**0.609**
Holidays	-	-	0.734	**0.825**
Barcelona	0.827	0.832	0.888	**0.900**

[1] The results are from [2].

Figure 4: Exemplar ranking result for *PGM* and *HPM*.



(a) BOF&Oxford (b) BOF&Holidays (c) BOF&Barcelona

(d) HE&Oxford (e) HE&Holidays (f) HE&Barcelona

Figure 5: mAP of *BOF*-based and *HE*-based systems against different sizes of image database with fixed reranking range.

thermore, as a flat and much simplified version of *HPM*, *BOF+1vs1+HV* (cf. *+HV* in Figure 5) can still achieve comparable performance. This is mainly because, in contrast to detecting conflicts at the visual word level in *HPM*, the proposed *1vs1* matching strategy operates at the point level, which makes it more accurate.

In addition, we observe that the improvement of *BOF+PGM* over *HPM* shrinks with the increasing scale of image collection. Due to the increasing number of distractor images in this case, the number of true-matching photos included in the (in this case fixed) reranking range is likely to decrease. However, within this range, it becomes increasingly easy to separate true matches from the false ones using spatial verification, with the consequence that all verification methods start performing similarly. As illustrated in Figure 6, the improvement achieved by PGM becomes significant again when we increase the reranking range with increasing image collection scale.

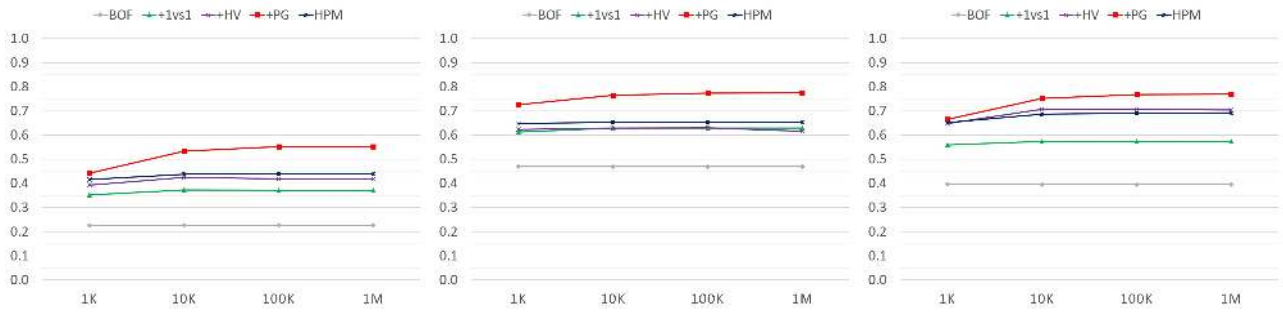
In the next experiment, we compare our best performing system variant, *HE+MA+PGM* with other state-of-the-art image retrieval systems in a similar setting: constructing the system on *generic* vocabulary, employing *multiple assignment*, using any form of spatial verification, and without query expansion. As summarized in Table 3, our system achieves state-of-the-art performance for image retrieval. The high performance achieved by [20, 19] on the *Oxford* dataset is mainly due to use of superior features, which can efficiently represent unrotated photos. This gain is, however, at the cost of worse performance for rotated photos, e.g., on the *Holiday* dataset. We note that we did not add query expansion [9, 8] and incremental spatial verification scheme [8] into our system, as they usually require re-calculating the correspondences for the new expanded query. We believe, however, that the proposed pairwise geometric matching method is compatible with these schemes.

Table 3: mAP comparison of different image retrieval system on generic vocabulary with spatial verification on top 200 (SP200) or top 1000 (SP1000) ranked photos.

	SP	Oxford	Holidays
Jégou *et al*. [13]	200	0.685	0.848
Philbin *et al*. [22]	200	0.598	-
HE+MA+PGM	200	**0.691**	**0.892**
Perd'och *et al*. [20]	1000	0.725	0.769
Mikulík *et al*. [19]	1000	**0.742**	0.749
HE+MA+PGM	1000	0.737	**0.892**

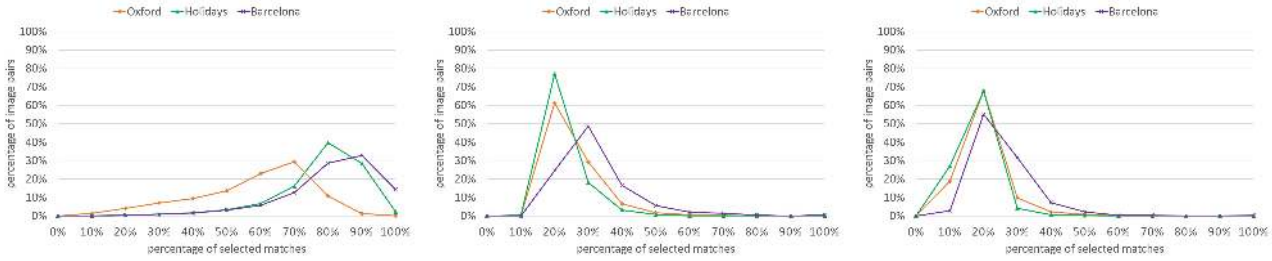
6.4. Run time efficiency

In the last experiment, we evaluate the run time efficiency of our system. To do this, we conduct spatial verification against all database images. We first analyze the ef-



(a) BOF&Oxford (b) BOF&Holidays (c) BOF&Barcelona

Figure 6: mAP of *BOF*-based system against 1M image database with different reranking ranges.



(a) 1vs1 (b) HV (c) 1vs1 and HV

Figure 7: Distribution of the percentage of selected matches after *1vs1* and *HV* steps, taken individually and together.

Table 4: Computing time and mAP comparison of *PGM* and *HPM* with spatial verification against all database images.

	Oxford		Holidays		Barcelona	
	Time[1]	mAP	Time[1]	mAP	Time[1]	mAP
PGM	**2.2**	**0.635**	**1.2**	**0.825**	1.1	**0.900**
HPM	2.8	0.527	1.7	0.734	**0.85**	0.888

[1] average matching time per pair of images in ms.

fect of the two filtering steps, *1vs1* amd *HV*, on reducing the size of the correspondence set. As illustrated in Figure 7, for about 60% of the image pairs, only 20% of matches remained to be checked after these two filtering steps, which dramatically reduces the influence of the pairwise operation on the overall run time. To evaluate the overall run time efficiency, we implement a toy version of our system in Java in a single-thread fashion to be comparable with the available binary code from *HPM*, and test it on a 2.3GHz 8-core processor. As summarized in Table 4, *PGM* achieves comparable run time efficiency, while significantly improving the performance. We also evaluate the query time of the entire retrieval system with spatial verification on top-1000 ranked images in the *BOF* setting. *PGM* achieves 2.7s, 1.6s and 0.7s for Oxford, Holidays and Barcelona datasets, respectively. In contrast, HPM consumes 2.9s, 2.7s and 0.7s.

7. Discussion

The results presented in the previous section indicate the suitability of the proposed pairwise geometric matching method as a solution for large-scale object retrieval at an acceptable computational cost. The superiority of *PGM* compared to the state-of-the-art solutions becomes evident in a context in which a high number of outliers in the initial correspondences generated by *BOF* and errors in detected features' scale, rotation and position hinder the fit of a specific model (e.g., RANSAC). *PGM* encodes not only scale and rotation information derived from the local points, but also their locations. This is achieved by using global scale and rotation relations to enforce the local consistency of geometric relations derived from the locations of pairwise correspondences. By mapping locations of points to pairwise rotation and scale, the approach is more tolerant to the detection noise. At the same time, using a number of filtering steps, *PGM* significantly reduces the number of correspondences that must be considered, which makes it possible for PGM to maintain high image matching reliability at a substantially reduced computational cost.

References

[1] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. In *Proc. CVPR '12*, 2012.

[2] Y. Avrithis and G. Tolias. Hough pyramid matching: Speeded-up geometry re-ranking for large scale image retrieval. *IJCV*, 107(1):1–19, 2014.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3):346–359, 2008.

[4] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang. Spatial-bag-of-features. In *Proc. CVPR '10*, 2010.

[5] G. Carneiro and A. Jepson. Flexible spatial configuration of local image features. *IEEE Trans. PAMI*, 29(12):2089–2104, 2007.

[6] G. Carneiro and A. D. Jepson. Flexible spatial models for grouping local image features. In *Proc. CVPR '04*, 2004.

[7] O. Chum, J. Matas, and S. Obdrzalek. Enhancing ransac by generalized model optimization. In *Proc. ACCV '04*, 2004.

[8] O. Chum, A. Mikulik, M. Perdoch, and J. Matas. Total recall ii: Query expansion revisited. In *Proc. CVPR '11*, 2011.

[9] 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 *Proc. ICCV '07*, 2007.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.

[11] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[12] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Proc. ECCV '08*, 2008.

[13] H. Jégou, M. Douze, and C. Schmid. On the burstiness of visual elements. In *Proc. CVPR '09*, 2009.

[14] H. Jégou, M. Douze, and C. Schmid. Improving bag-of-features for large scale image search. *IJCV*, 87(3):316–336, 2010.

[15] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[16] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *Proc. ICCV '05*, volume 2, pages 1482–1489 Vol. 2, 2005.

[17] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[18] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[19] A. Mikulík, M. Perdoch, O. Chum, and J. Matas. Learning a fine vocabulary. In *Proc. ECCV '10*. 2010.

[20] M. Perd'och, O. Chum, and J. Matas. Efficient representation of local geometry for large scale object retrieval. In *Proc. CVPR '09*, 2009.

[21] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR '07*, 2007.

[22] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Proc. CVPR '08*, 2008.

[23] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *Proc. CVPR '12*, 2012.

[24] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. In *Proc. ICCV '03*, 2003.

[25] G. Tolias and Y. Avrithis. Speeded-up, relaxed spatial matching. In *Proc. ICCV '11*, 2011.

[26] Z. Wu, Q. Ke, M. Isard, and J. Sun. Bundling features for large scale partial-duplicate web image search. In *Proc. CVPR '09*, 2009.

[27] L. Yang, B. Geng, Y. Cai, A. Hanjalic, and X.-S. Hua. Object retrieval using visual query context. *IEEE Trans. Multimedia*, 13(6):1295–1307, 2011.

[28] Y. Zhang, Z. Jia, and T. Chen. Image retrieval with geometry-preserving visual phrases. In *Proc. CVPR '11*, 2011.