

Enhancing instance search with weak geometric correlation consistency

Zhenxing Zhang^a, Rami Albatal^a, Cathal Gurrin^a, Alan F. Smeaton^a

^a*Insight Centre for Data Analytics
School of Computing, Dublin City University
Glasnevin, Dublin 9, Ireland*

5

Abstract

Finding object instances from within in large image collections is a challenging problem with many practical applications. Recent methods inspired by text retrieval have achieved good results, however a re-ranking stage based on spatial verification may still be required to boost performance. To improve the effectiveness of such instance retrieval systems while avoiding the computational complexity of a re-ranking stage, we explore the geometric correlations among local features, and we incorporate these correlations with each individual match to form a transformation consistency in rotation and scale space. This weak geometric correlation consistency can be used to effectively eliminate inconsistent feature matches in instance retrieval and can be applied to all candidate images at a low computational cost.

15

Experimental results on three standard evaluation benchmarks show that the proposed approach results in a substantial performance improvement when compared with other state-of-the-art methods. In addition, the evaluation results from participating in the Instance Search Task in the TRECVid evaluation campaign also suggest that our proposed approach enhances retrieval performance for large scale video collections.

Keywords:

20

Multimedia Indexing, Information retrieval, Instance Search, Weak geometric correlation consistency

Email addresses: zzhang@computing.dcu.ie (Zhenxing Zhang), ramibatal@gmail.com (Rami Albatal), cgurrin@computing.dcu.ie (Cathal Gurrin), alan.smeaton@dcu.ie (Alan Smeaton)
Preprint submitted to Neural Computing December 5, 2016

1. Introduction

Given a query image of an object, the objective of this work is to find images that contain recognisable instances of the object from a large image collection, henceforth referred to as “instance search”. A successful application of instance search requires efficient retrieval of instance images with high accuracy, possibly under varying imaging conditions such as rotation, viewpoint, zoom level, occlusion, and so on.

Instance search is an interesting, yet challenging, problem and has attracted significant research attention in recent years. Most of the state-of-the-art approaches [11], [16], [19] have been developed based on the Bag-of-Visual-Words (BoVW) representation first introduced by J. Sivic et al. [3]. This representation framework successfully made use of the discriminative power of local feature descriptors, for example SIFT [1] and SURF [2] which are generally robust to multiple changes in imaging conditions, and are applied to build a statistical representations for each image in the database. At query time, the BoVW representation may take advantage of indexing techniques such as inverted files [4] to provide fast retrieval over large collections.

However this representation leads to a loss of the ability to encode spatial information between local features, so spatial verification [16] was subsequently introduced to improve retrieval accuracy. Based on the observation that the layout of local features from the query object and its instances should share the same or similar geometric structure, there could be only one feature correspondence for any given feature in the query object. So the geometric layout of query objects was adopted to verify the spatial consistency between initial matched local features. Generally, such spatial verification algorithms [11] [16] were applied to train models to capture the transformation in spatial space (normally 3-D transformation including position, rotation, scale changes) between feature correspondences and to fit them to the initial correspondences to eliminate inconsistent matches in order to refine the results. However these techniques such as RANSAC [25] are normally computationally expensive, and

so in practice they are only be applied as a post-processing step to the a limited set of top-ranked images in the initial results.

In our work, we address the challenges of improving the efficiency and robustness of examining the consistency between local feature matches to enhance the retrieval performance of instance search systems. However, in contrast to previous spatial verification technologies, we followed the work proposed by H. Jégou et al. [5] to efficiently apply spatial verification. Instead of estimating full spatial transformation models, which is what H. Jégou et al. [5] propose, we focused on building weak geometric constraints in 2-D space, specifically in the rotation and scale parameters, to examine each individual feature correspondences. This leads to a much reduced computational expense and although having only two spatial parameters are not sufficient to map objects from one image to another, the weak geometric constraints could help us to filter out inconsistent feature correspondences at a very low computational cost. The real consequence of this is that reduced cost makes this suitable for very large data collections and this is an enhancement on the original work by H. Jégou et al. [5]. In the work from H. Jégou et al. [5], we observe that their approach considered feature matches independently and ignored the geometric correlation between local features, and thus it performed less effectively when performing search in more challenging datasets like FlickrLogos-27 [18]. In our work, we believe that the geometric correlation between reliable feature matches should also be consistent to the weak geometric constraints, just like each individual feature match. Based on that, we propose a scheme to incorporate the geometric correlations between matched feature correspondences to form a weak geometric correlation consistency to improve the effectiveness of spatial verification. Thus the main contribution of the paper is the construction of weak geometric constraints in 2-D space allowing this to be done with low computational cost while still supporting geometric correlations but in a much more efficient way when scaled to large datasets and we demonstrate this on some large data collection

This paper is organised as follows. In the next section we present some of the most relevant work in the area and following that, in Section 3 we introduce the

idea of weak geometric correlation consistency. In Section 4 we present a brief motivation for the need for improvement in instance search matching, and our experiments are introduced and presented in Section 5. Our experimental results on the TRECVid INS task are presented in Section 5. In summary, the paper illustrates that our proposed method is more reliable, and also more tractable for large image collections, and leads to an overall significant improvement of instance search performance compared to state-of-the-art methods.

2. Related work

In this section, we briefly review the development of visual instance retrieval systems and discuss existing approaches to improving retrieval performance by using the geometric information.

The idea of re-ranking a visual search and carrying out some form of dimensionality reduction in the process, is now receiving some attention in the research community. One approach [29] is based on exploiting the overall manifold structure of the whole of the dataset and preserving relationships between example images which have been labelled, effectively mapping a data set to a lower dimensionality under constraints of preservation.

[3] were the first to address instance search using a BoVW representation combined with scalable textual retrieval techniques. Subsequently, a number of techniques have been proposed to improve performance. The work reported in [11] suggested using a very high dimensional vocabulary (1 million visual words) during the quantization process. This method improved the retrieval precision with more discriminative visual words, and also increased retrieval efficiency with more sparse image representations, especially for large scale database. [13] brought query expansion techniques to the visual search domain and improved instance recall by expanding the query information. For further improvement on the retrieval performance, both approaches added the spatial verification stage to re-rank their results in order to remove noisy or ambiguous visual words.

While geometric-based constraints are of ultra-importance for some image-matching applications like fingerprint matching, as described in [28], recent work reported in [7], [10], [9] and [8] extended the BoVW approach by encoding
115 the geometric information around the local features into the representation and refine the matching based on visual words. Those methods were very sensitive to the change in imaging condition and made them only suitable for partial-duplicate image search.

For many high-dimensional data processing applications which involve image data like search, clustering, duplicate detection, etc., we often have too many
120 features in the image representation, many of which are often redundant and noisy. Feature selection is one technique for dimensionality reduction that involves identifying a subset of the most useful features. In [30], a novel unsupervised feature selection algorithm, named clustering-guided sparse structural
125 learning (CGSSL), is proposed that, like our work, helps to discriminate among features, in our case using geometric constraints.

One of the other ways to reduce the feature space is to learn a subspace, at retrieval time, and to perform the matching within that subspace. In [], the authors proposed exactly that, and to guarantee their subspace to be compact
130 and discriminative, the intrinsic geometric structure of data, and the local and global structural consistencies over labels were exploited simultaneously in a similar way to the way we exploit geometric constraints.

Even more recently, alternative approaches have been developed to implicitly verify the feature matches with respect to the consistency of their geometric
135 relations, i.e., scaling, orientation, and location, in the Hough transformation space. [12] developed a linear algorithm to effectively compute pairwise affinities of correspondences in a 4-dimensional transformation space by applying a pyramid matching model constructed from each single feature correspondence. [5], increased the reliability of feature matches against imaging condition changes
140 by applying weak constraints to verify the scaling and orientation relations consistency according to the dominant transformation found in the transformation space. Similarly, [15] proposed to represent the feature points' geometric in-

formation using topology-base graphs and verified the spatial consistency by performing a graph matching.

145 Finally, if we broaden out the function from just instance search to other content-based operations we find that there are other examples of work which uses some form of spatial coding, to encode the spatial relationships among local features in an image. Work described in [10] shows significant improvement in performance, but accuracy and computational cost, for the task of detecting
150 near-duplicate images in web search.

Our proposed method follows the direction of implicitly verifying feature matches do exist in an instance search in order to reduce the computational cost. However compared to existing work, which focused on individual correspondences, our proposed method also considers the spatial consistency for the
155 geometric correlations between matched feature correspondences, while maintaining the efficiency and increasing the effectiveness of the instance search systems.

3. Weak Geometric Correlation Consistency

In the BoVW representation for images, local features are first extracted
160 from each image to encode invariant visual information into feature vectors. Generally, a feature vector is defined as $\vec{v}(x, y, \theta, \sigma, q)$, where variables $\{x, y, \theta, \sigma\}$ stand for the local salient point’s 2-D spatial location, dominant orientation, and most stable scale, respectively while q represents a 128-D feature vector to describe the local region. For a query image I_q and candidate image I_c , a set of
165 initial matching features $C_{initial}$ could be established by examining the feature vector q . The task of spatial verification is to eliminate the unreliable feature matches and only retain the matches set C_{stable} that link the patches of the same object. The following equation formulates this process:

$$C_{stable} = \{m_i \in C_{initial} \quad \text{and} \quad f_{sp}(m_i) = 1\} \quad (1)$$

where m_i stands for the i^{th} feature match in the initial match set and f_{sp} stands
170 for the spatial verification function for assessing a geometric consistency. Take

the weak geometric consistency of [5] for example, the verification function in their work could be expressed as follows:

$$f_{sp} = \begin{cases} 1 & \text{if } \Delta\theta \in D_\theta \text{ and } \Delta\sigma \in D_\sigma \\ 0 & \text{if otherwise} \end{cases} \quad (2)$$

where $\Delta\theta$ and $\Delta\sigma$ is the geometric transformation for an individual feature match and D_θ and D_σ is the weak transformation constraints in rotation and scale space, which were the approximation of the full spatial transformation. Experimentally, the value of these constraints was calculated from the initial local feature correspondences.

3.1. Motivation

We take the geometric correlation among local features into consideration and hypothesize that the pairwise geometric correlation between consistent matches should also be consistent and should follow the same spatial transformation between objects. So instead of verifying the geometric consistency for each match individually, we propose a novel approach to verify the consistency between pairwise geometric correlations along with their corresponding feature points. So for a given pair of feature matches m_l and m_n , we define the proposed spatial verification function as following:

$$f_{sp} = \begin{cases} 1 & \text{if } \Delta\theta, \Delta\theta_{l \rightarrow n} \in D_\theta \text{ and } \Delta\sigma, \Delta\sigma_{l \rightarrow n} \in D_\sigma \\ 0 & \text{if otherwise} \end{cases} \quad (3)$$

where $\Delta\theta_{l \rightarrow n}$ and $\Delta\sigma_{l \rightarrow n}$ represents spatial transformation of the geometric correlation from feature match m_l to m_n .

We call our proposed approach Weak Geometric Correlation Consistency (WGCC), and Figure 1 demonstrates our idea of using geometric correlations to assess the reliability of feature matches. The object of interest (the front cover of a box) is highlighted with a dark yellow box. To begin with, we have three initial feature matches for spatial validation. Matches (A, A') , (B, B') are considered to be consistent because the spatial transformation (i.e. translation,

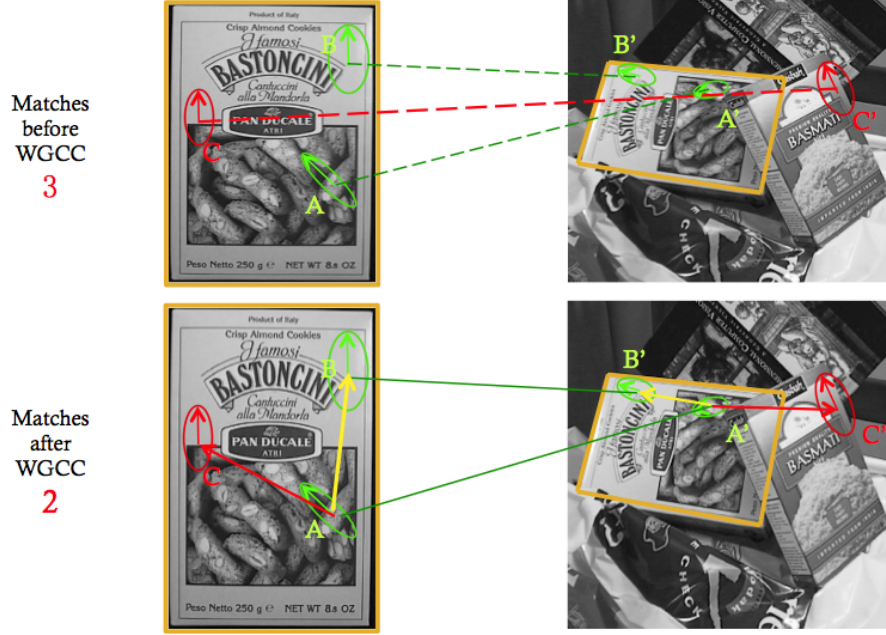


Figure 1: An illustration of verifying consistency of feature matches using geometric correlations. The green(red) line indicates the consistent(inconsistent) feature matches.

195 rotation and scale change) is consistent between (A, A') , (B, B') and their correlation $(AB, A'B')$. On the other hand, match (C, C') is filtered out due to the fact that geometric correlation between $(AC, A'C')$ is not consistent with the spatial transformation. Hence, we can successfully eliminate the inconsistent feature matches despite the fact that they may obey weak spatial constraints
200 individually.

3.2. Implementation

The explicit examination of all correlations between initial feature matches is a non-trivial task. If we take a total number of N initial matches as an example, the potential pairwise correlation could be modeled as $O(C_N^2)$. The
205 initial feature matching number N is usually large in practical systems, and this will result in a high computational cost to verify all the correlations, and thus will make the solution less attractive for large-scale image collections.

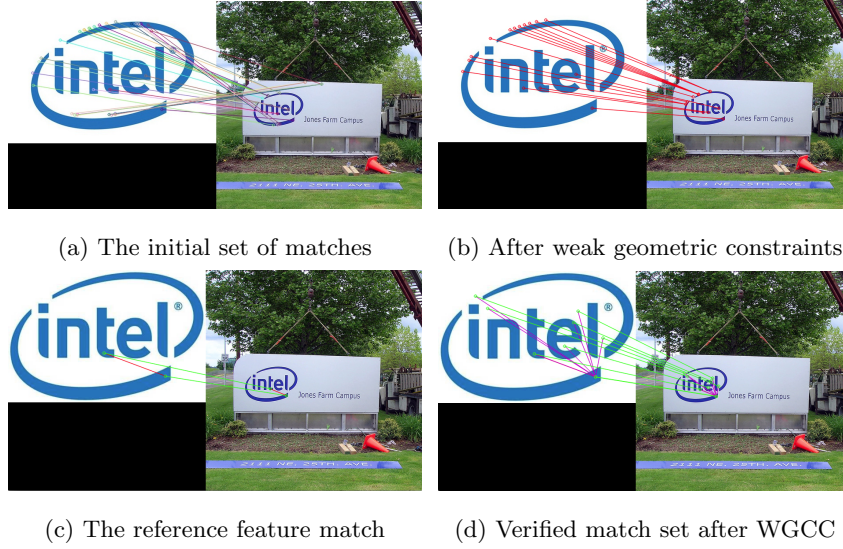


Figure 2: An illustration of applying WGCC on the initial set of feature matches to obtain the consistent feature matches.

In this work, we propose a three-step process to reduce the complexity of verifying the geometric correlation consistency, and to make it applicable at low-cost for large-scale instance search systems, which is our key goal. The key idea is to obtain a feature match as a reference point between the initial set of feature matches and then examine only the $O(N)$ correlations between each match and the reference match. These three steps are described in the following paragraphs and an example output for each step is shown in Figure 2.

Estimating weak geometric constraints. To begin with, we establish a weak geometric transformation, specifically rotation and scaling, in the spatial space from the initial set of feature matches. The transformation parameters, rotation angle $\Delta\theta$ and scaling factor $\Delta\sigma$ for each feature match were denoted as:

$$\Delta\theta = \theta_m - \theta_i, \Delta\sigma = \sigma_m / \sigma_i \quad (4)$$

In order to reduce the sensitivity to non-rigid deformation [27], we quantize the value of the parameters into bins to estimate an approximated transformation. We use a factor of 30 degrees to divide the rotation range of 360 degrees into

12 bins, and a factor of 0.5 to divide the scale range between 0 to 4 into 8 bins. To avoid any possible bin quantization error, each feature match votes to the closest two bins in each parameter space. The Hough voting scheme [24] was applied in searching of the dominant value D_θ and D_σ to form weak geometric constraints for two purposes. Firstly, we can reduce the computational complexity of following process by eliminating the matches that do not obey the constraints. Secondly, these weak constraints will be used to assess the transformation consistency for geometric correlation to obtain the reliable matches.

Identifying the reference matching correspondence. In this step, we aim to determine the strongest feature matches which will be served as a reference match in the verifying geometric correlations step. We follow the approach of [15] and adopt a topology-based graph match for this purpose. To represent the topology structure for objects, we create a delaunay triangulation mesh from the geometric layout among the feature points in object plane. We then find the strongest feature matches which correspond to the common edges between topology graphs by performing a graph matching.

Verifying weak consistency for geometric correlations. The final step is focused on identifying the most reliable feature matches by verifying the consistency of the geometric correlations from each feature match to the reference match. Suppose we have a feature match m_l and a reference match m_n between image Q and D , the geometric correlation from m_l to m_n in image Q could be expressed as a vector $v_{l \rightarrow n} = (x_l, y_l) - (x_n, y_n)$ where x, y represent the 2D location of corresponding feature points in image Q for match m_l and m_n respectively. Similarly we can express the geometric correlation between m_l and m_n in image D as a vector $v'_{l \rightarrow n} = (x'_l, y'_l) - (x'_n, y'_n)$. Then the transformation parameters in orientation $\Delta\theta_{l \rightarrow n}$ and the scale $\Delta\sigma_{l \rightarrow n}$ between geometric correlations can be defined as:

$$\Delta\theta_{l \rightarrow n} = \arccos \frac{\|v_{l \rightarrow n}\| \|v'_{l \rightarrow n}\|}{v_{l \rightarrow n} \cdot v'_{l \rightarrow n}}, \quad \Delta\sigma_{l \rightarrow n} = \frac{\|v'_{l \rightarrow n}\|}{\|v_{l \rightarrow n}\|} \quad (5)$$

It is now possible to assess spatial consistency by verifying the transformation parameters values with weak constraints according to equation 3 And as illus-

Figure 3: Overall framework of our WGCC approach in instance search system



trated in step (d) in figure 2, final set of reliable feature matches can be obtained with further filter out inconsistent matches.

The high-level components of our instance search systems are outlined in figure 3. This diagram illustrates an overview of instance search pipeline from which we obtain our initial results and apply our proposed approach for re-ranking to obtain an improved search results.

3.3. Computational Complexity

The major computational cost in the proposed scheme is in the second step where we build the triangulation mesh and discover the reference matches by identifying the common edges. These computations are closely related to the total number of feature matches. Since we have already built weak geometric constraints in the first step to verify the initial feature matches, only a subset of smaller sets of feature matches (the cardinality of this set is denoted by n) needs to be conducted in this step, which leads to a cost of $O(n \log n)$. In the end, $O(n)$ operations are required to perform the geometric correlation verification which is much less than $O(C_n^2)$ required for a full verification of all the possible geometric correlations, and this is the basis for the saving in computational cost which allows our proposed method to be scaled to large image and video collections.

4. Understanding Performance Improvement in Object Matching

In this section, we try to understand the performance improvement provided by our proposed approach by locating a specific scene or object in a digital

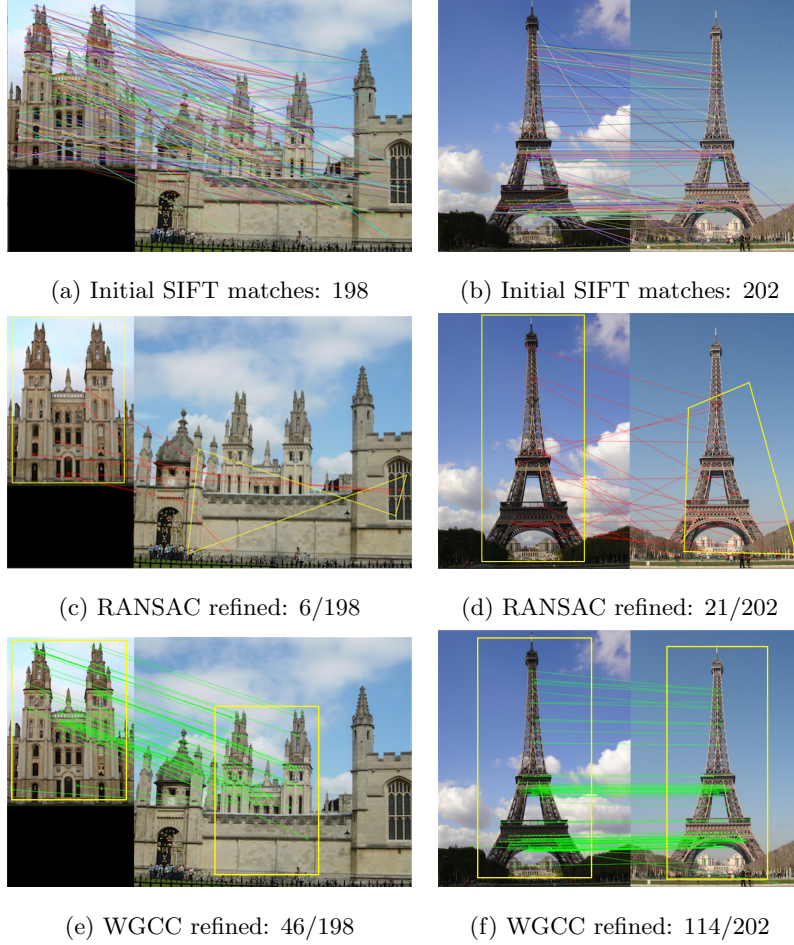


Figure 4: Comparing original object matching results (a) and (b) with the RANSAC algorithm (c) and (d) and our proposed WGCC algorithm (e) and (f), in two object matching examples. After applying our WGCC algorithm, many more consistent matches are found and the object is better localized in the target image.

image. As discussed in section 3, this work aims to refine the initial match-
 275 ing points between two images by verifying the spatial correlation to eliminate
 outliers. So, we compared the classic RANSAC based approach [25] with our
 proposed WGCC approach in perform object matching tasks between a pair of
 images. We expect that our approach will outperform the RANSAC algorithm
 in terms of efficiency since the geometric correlation consistency provided by

280 our algorithm could be used to quickly and robustly filter out any outliers.

Figure 4 demonstrates two object matching examples with the RANSAC algorithm and our proposed WGCC algorithm. In this figure, the query object is shown on the left side of each matching pair, and the target image is shown on the right. The first row presents the initial matching from calculating the distance between the SIFT descriptors, the second row presents the refined matching after applying the RANSAC algorithm, and the last row presents the consistent matching after applying our WGCC algorithm. A comparison of the two object matching examples clearly shows that our proposed algorithm increases the number of consistent matches and has successfully identified object localization in the target images. To quantitatively measure the effectiveness of our WGCC approach, we selected 10 pairs of images to perform the object matching task from the Oxford building test collection and calculated the correct ratio of reliable matching from initial feature correspondences. Our WGCC method performed about 5 times better than the standard RANSAC method in average for preserving the consistent feature correspondences. After applying WGCC method into instance search tasks, we expected our approach could help to reduce object matching confusion and boost the retrieval performance by refining each image in the initial rank list.

5. Instance Search Experiments

300 The goal of our experiments is to assess the performance of the proposed weak geometric correlation consistency methods in instance search tasks. In order to achieve this, a complete instance search system was developed and comparative experiments were designed to evaluate the retrieval performance against state-of-the-art approaches on three standard and publicly-available benchmarks.

305 In the rest of this section, we introduce the three chosen benchmark datasets, describe the evaluation protocol and analyse the experimental results by comparing them to three state-of-the-art approaches described in [11], [5] and [15]

respectively.

310 5.1. Datasets

The datasets chosen were the Oxford, Pairs6K and FlickrLogis-32 datasets. Each of these datasets includes a set of queries and relevance judgements.

The Oxford dataset. This dataset [11] contains 5,062 high resolution images crawled from Flickr using texture queries for pictures of famous Oxford
315 landmarks. 11 building topics with 55 images as queries was provided with manually annotated ground truth for users to evaluate retrieval performance. The images are considered to be positive if more than 25% of the instance is clearly visible in the image frame.

Pairs6K This collection [16] consists of 6,412 images collected by searching
320 for particular Paris landmarks from Flickr. In total, 11 Landmarks with 55 image queries was provided with manually annotated ground truth for users to evaluate retrieval performance. The images are considered to be positive if more than 25% of the instance is clearly visible in the image frame.

FlickrLogos-27 This dataset [18] consists of 5,107 images including 810
325 annotated positive images corresponding to 27 classes of commercial brand logos and 4,207 distraction images that depict its own logo class. This is a very challenging dataset because the positive images share much more visually similar regions with the distraction images and have more noisy background. For each logo, 5 query example images are given for evaluation purposes.

330 5.2. Evaluation protocol

A standard evaluation protocol based on the classic BoVW scheme was adopted to assess the improvements of our proposed method for instance search. The Hessian detector and SURF descriptor implemented in the OpenCV Library [20] were used to extract the local features from database images. Subse-
335 quently, a visual vocabulary was generated using the approximate K-means algorithm described by [11] to quantize each feature into visual words for indexing. After that, the represented visual words (along with auxiliary information, e.g.

the geometric information) are indexed in an inverted structure for the retrieval process. When performing the search tasks, the candidate images sharing same
340 visual words are retrieved from the database collections. Auxiliary information is used to perform the spatial verification used to improve retrieval performance.

We measured the mean Average Precision (mAP) score of the top 1,000 results to evaluate retrieval accuracy. mAP is defined as the mean of the average precision (AP) over all queries. To evaluate retrieval efficiency, we also record
345 the response time accurate to one hundredths of a second. Each approach was implemented and evaluated on the same computing hardware.

The experiments were carried out on a desktop computer with 4-core 2.3 GHz CPU and 8G RAM. Only one core was used when performing the task.

5.3. Approaches for Comparison

350 The Weak Geometric Correlation Consistency (WGCC) was compared against the standard BoVW approach as the baseline, but it is also compared against two other advanced approaches; Weak Geometric Consistency (WGC) and Delaunay Triangulation(DT). We choose this two advanced methods and compare their performance to demonstrate the effectiveness of our proposed approach
355 due to two main reasons. Firstly, these two approaches both employ the geometric information among local features to enhance the performance of baseline method. Secondly, we aim to verify the effectiveness of every step in our proposed three-step algorithm, where weak geometric consistency [5] and delaunay triangulation algorithm [15] are chose to measure the effectiveness of step one
360 and two respectively.

BoF [11]. The baseline approach was based on [11] with a vocabulary of 1M words which had been shown to give the best performance. The $tf \times idf$ weighting scheme and hard assignment was used to keep a consistent setting for all system implementations.

365 **WGC** [5]. We chose this approach for evaluation because this method assessed each feature match by verifying its transformation against a weak geometric consistency to increase the robustness in changing of rotation and scale

space. The constraints for geometric consistency was obtained by converting the parameter values into a Hough transformation space.

370 **DT [15]**. This approach makes use of relations between matched points in a 2-dimensional translation space to improve the matching reliability between two sets of features. It used the Delaunay Triangulation (DT) based graph representation to model and match the layout topology of initial matched feature points. A hamming embedding signature was used to enforce point-to-point
375 matches and to ensure the number of nodes in each graph is identical.

WGCC. This is our proposed method and the contribution of this work, as described in section 3. We follow the recent work in feature search in high-dimensional spaces and use the product-quantization based algorithm [6] to build up search components for initial feature matching. Then we applied the
380 proposed weak geometric correlation consistency (WGCC) for spatial verification, and reduction of computational cost.

5.4. Results and discussion

In this section, we present and discuss the quantitative results from our evaluation experiments and we demonstrate the performance improvement provided
385 by our proposed approach in comparison with other advanced approaches.

5.4.1. Comparison Using mean Average Precision(mAP)

Table 1 presents the experimental results of comparing our proposed approach WGCC with the baseline and two enhanced approaches, on the three benchmark datasets. We observe that the advanced approaches for spatial verification consistently improves performance in terms of mAP compared to the
390 baseline. Compared to the other two advanced systems, our proposed approach achieves the best results. This is clearst on the FlickrLogos dataset, where our approach has a 59% relative improvement in the mAP performance from the baseline’s 0.145 to 0.231 in our method. This proves that our approach is strong
395 enough to reject inconsistent feature matches, while also being flexible enough to keep the evidence from locally consistent patches.

Table 1: Comparison in mAP score and response time between our proposed WGCC and the baseline and two other state-of-the-art advanced approaches, on Oxford, Pairs6K and FlickrLogo-27 datasets.

	Oxford		Pairs6K		FlickrLogos	
Methods	mAP	Time ¹	mAP	Time ¹	mAP	Time ¹
BoF	0.489	0.46	0.526	0.62	0.145	0.26
WGC	0.530	1.06	0.576	1.12	0.193	0.41
DT	0.542	0.86	0.546	0.89	0.201	0.31
WGCC	0.693	1.07	0.607	1.23	0.231	1.06

Time¹ measures the average response time per query in second, excluding feature extraction.

5.4.2. Comparison Using Precision-Recall Curves

Figure 5 shows some examples of the improvement in mAP obtained by the proposed WGCC approach compared to the baseline system in terms of precision recall curves. The object of interest is delimited in the yellow box from the query image on left side of each sub-figure. The Precision-Recall curve is displayed on the right side with baseline results shown in a blue line and the WGCC method shown in red. The gap area between two lines indicates the performance improvement for our methods. We could understand that our proposed method refined the initial results by successfully re-ranking the relevant images in high order.

5.4.3. Comparison Using Top-10 Results

In order to demonstrate the performance gained with our proposed approach, we also compared the top-10 results for three query topics returned by the baseline approach and our WGCC approach in Figure 6. The left column on the figure presents the three query objects from three test collections. On the right-hand side, the top-10 results are displayed from left to right. For each query topic, the upper rows are the results from the baseline system and the lower

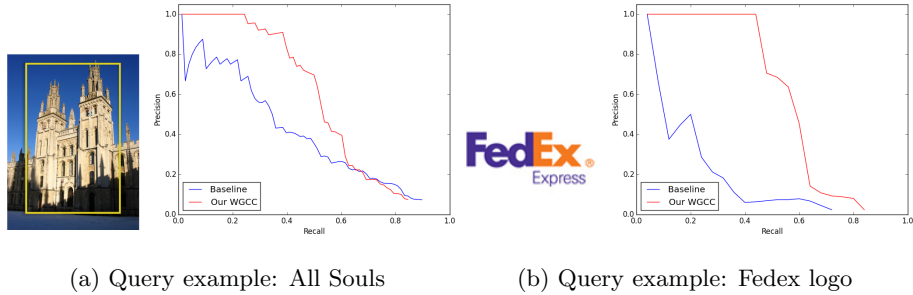


Figure 5: Performance improvement in the precision-recall curve obtained by using proposed approaches compared to the baseline system.

rows are the results from our WGCC method. The non-relevant result images
 415 are outlined with red surrounding boxes and the remainder are relevant results.
 As illustrated in the figure, our WGCC method is especially good at returning
 relevant results in the top-10 ranking. Since our WGCC approach preserved
 the geometric layout of object instances with weak geometric constraints, so
 our approach would rank the irrelevant images to a much lower place even if
 420 they have many initial matching features with the query image.

5.4.4. Comparison Using Efficiency

At run-time, our proposed method achieved comparable retrieval efficiency
 with the two advanced approaches while providing better accuracy. Table 1
 also illustrates the average response time over all the query topics for each test
 425 collection. Although the retrieval efficiency of WGCC was slightly less than the
 other approaches, retrieval efficiency could be optimised by adopting parallel
 computing approaches.

6. Large Scale Instance Search Experiments

To evaluate the performance of our proposed approach on large-scale in-
 430 stance search, we participated the instance search task (INS) in the annual
 TRECVID Retrieval Evaluation workshop [21, 22].



(a) Query example: 'All Soul' from the Oxford building dataset



(b) Query example: 'Ferrari Logo' from the Flickr Logos 27 dataset



(c) Query example: 'Eiffel Tower' from the Paris6K dataset

Figure 6: Comparison of Top-10 results in ranking lists. The image on the left column is the query images, and on the right-hand side, the top-10 results are displayed from left to right. For each query, the upper rows are the results from baseline system and the lower rows are the results from our WGCC method. The red box indicate the irrelevant result images.

6.1. INS in TRECVID 2014 and 2015

TRECVID is an international video search campaign where researchers evaluate and demonstrate the efficiency and effectiveness of their video retrieval approaches on a shared test collection via an open, metrics-based evaluation. The instance search task specially focused on finding video segments which contains a certain specific person, object, or landmark, given a visual query image. During the event in 2015, 30 topics with ground-truth data were given for each participant to find the top 1,000 shots which are most likely to contain a recog-

440 nizable instance from a large collection of about 200 hours of video. To evaluate the retrieval performance for each submission, the standard evaluation metrics such as mean Average Precision (mAP), Precision at K (P@5, P@10, ...) was computed.

6.2. Our Participation and Results Discussion

445 **Search Engine** We built a search engine [26] which will take an image as a search query and then sort a keyframe taken from every video shot in the search collection according to the likelihood that it may contain an instance of the topic object. The Hessian detector and SIFT descriptor implemented in the OpenCV Library [20] were used to extract the local features from database
450 images. Subsequently, a visual vocabulary with 1 million visual words was generated using the approximate K-means algorithm [11] to quantize each feature into a visual word to build the bag of feature representation. We adopted the standard $tf \times idf$ weighting algorithm [23] to calculate the discriminative power of each visual word, and then computed the relevance score for each image in
455 the collection.

Our submissions and Results In order to reveal the efficiency of our WGCC approach for large scale instance search tasks, we submitted two automatic runs with different settings to the INS task in TRECVID in 2015. In the first run, the initial result was refined using the classic RANSAC algorithm as
460 the spatial verification step, and the second run was conducted with our proposed WGCC algorithm. For each run, all the 30 search queries were performed and the top-1000 results in the ranking list were submitted for evaluation.

Since all the instance search tasks were completed on the same dataset, we can directly compare the results from the two runs to demonstrate the efficiency
465 of their retrieval performance. Figure 7 illustrates the sorted list of mAP scores for all the tasks from the two runs in our TRECVID 2015 experiment. The red columns stands for the scores obtained with our WGCC approach in the second run and the blue columns are the scores from the RANSAC algorithm in the first run.

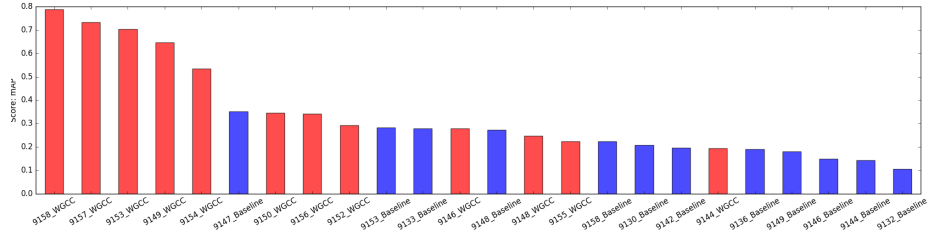


Figure 7: Comparison in mAP scores between our submissions in INS tasks in TRECvid 2015. The red columns are scores from our WGCC algorithm and blue columns are scores from the RANSAC algorithm

The figure shows that our WGCC approach significantly outperformed the RANSAC algorithm in 5 search tasks. And the best mAP score for our WGCC approach achieved 0.79, in contrast only 0.34 for the RANSAC algorithm. This experiment strongly proves that our proposed WGCC algorithm successfully improves the performance of instance search on very large data collections.

7. Conclusion

This paper proposed a novel approach to improve retrieval performance of instance search systems by combining pairwise geometric correlations with individual feature match transformation in order to form a weak geometric correlation consistency. This model eliminates inconsistent feature matches while keeping reliable matches using locally spatial correlations. Our experiments show that our approach consistently outperforms our baseline system in three standard benchmark evaluations and achieves improved results when compared with two advanced systems. This indicates the effectiveness of our method for spatial verification. Another positive aspect of our experimental results is that other advanced technologies, such as automatic query expansion [13], or re-ranking based on full spatial verification [11] are compatible with our proposed method and could be used as complementary components to further improve retrieval performance. In future work, we will investigate how to incorporate WGCC methods in vary large data collections, e.g. collection with millions of images.

8. Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under grant number SFI/12/RC/2289, as well as the financial support of the Norwegian Research Council's iAD project under grant number 174867.

- [1] D. Lowe. Distinctive image features from scale-invariant key points. IJCV, 60(2): 91-110, 2004.
- [2] H. Bay , A. Ess , T. Tuytelaars , L. Van Gool. Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding, v.110 n.3, p.346-359, June, 2008.
- [3] J. Sivic, A. Zisserman. Video Google: Efficient Visual Search of Videos in Toward Category-Level Object Recognition, Springer, Volume 4170, page 127–144, 2006
- [4] J. Zobel, A. Moffat, K. Ramamohanarao, Inverted files versus signature files for text indexing, ACM Trans. Database Systems 23, 453490, 1998.
- [5] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In ECCV, October 2008.
- [6] H. Jégou, M. Douze and C. Schmid. Product quantization for nearest neighbor search in IEEE Transaction on Pattern Analysis and Machine Intelligence, 2011
- [7] Z. Wu, Q. Ke, M. Isard, J. Sun. Bundling features for large scale partial-duplicate web image search in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.
- [8] R. Albatal, P. Mulhem, and Y. Chiaramella. Visual phrases for automatic images annotation in Proceedings of CBMI, 2010.
- [9] S. Romberg, R. Lienhart, Bundle min-hashing for logo recognition, in Proceedings of International Conference of Multimedia Retrieval, 2013

- [10] W. Zhou, Y. Lu, H. Li, Y. Song, Q. Tian. Spatial coding for large scale partial-duplicate web image search. in Proceedings of the ACM conference in Multimedia 2010.
- [11] J. Philbin , O. Chum, M. Isard, J. Sivic, and A. Zisserman, Object retrieval with large vocabularies and fast spatial matching in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007
- [12] Y. Avrithis and G. Tolias. Hough pyramid matching: Speeded-up geometry re-ranking for large scale image retrieval. IJCV, 107(1):119, 2014.
- [13] O. Chum, J. Philbin, J. Sivic, M. Isard, A. Zisserman. Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval in IEEE International Conference on Computer Vision, 2007
- [14] Y. Zhang, Z.Jia, and T. Chen, Image retrieval with geometry-preserving visual phrases in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2011.
- [15] W. Zhang and C.-W. Ngo, Searching visual instances with topology checking and context modeling, in Proceedings of ICMR, 2013.
- [16] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008
- [17] J. Revaud, M. Douze, C. Schmid. Correlation-Based Burstiness for Logo Retrieval in Preceeding of the ACM International Conference on Multimedia, Oct 2012
- [18] S. Romberg, L. G. Pueyo, R. Lienhart, R. van Zwol, Scalable Logo Recognition in Real-World Images in ACM International Conference on Multimedia Retrieval 2011 (ICMR11), Trento, April 2011.

- [19] Y. Kalantidis, LG. Pueyo, M. Trevisiol, R. van Zwol, Y. Avrithis. Scalable
545 Triangulation-based Logo Recognition. In Proceedings of ACM International Conference on Multimedia Retrieval (ICMR 2011), Trento, Italy, April 2011.
- [20] G. Bradski, The OpenCV Library in Dr. Dobbs Journal of Software Tools, 2000
- 550 [21] Over, P., Awad, G., Michel, M., Fiscus, J., Kraaij, W., Smeaton, A. F., Quenot, G., and Ordelman, R. (2015). TRECVID 2015 an overview of the goals, tasks, data, evaluation mechanisms and metrics. In Proceedings of TRECVID 2015.
- [22] Smeaton, A.F., Over, P. and Kraaij, W. (2006). Evaluation campaigns and
555 TRECVID. Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pages 321–330, Santa Barbara, California, USA.
- [23] Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press, New York, NY, USA.
- 560 [24] Shapiro, L. G. and Stockman, G. C. (2001). Computer Vision. Prentice Hall, New Jersey, USA.
- [25] Fischler, M. A. and Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Commun. ACM, New York, NY, USA.
- 565 [26] K. McGuinness, Mohedano, E., Salvador, A., Zhang, Z. X., Marsden, M., Wang, P., Jargalsaikhan, I., Antony, J., Gir-i-Nieto, X., Satoh, S. 'ichi, O'Connor, N., and Smeaton, A. F., Insight DCU at TRECVID 2015, in TRECVID 2015 Workshop, Gaithersburg, MD, USA, 2015.
- [27] L. Zagorchev and A. Goshtasby, "A comparative study of transformation
570 functions for nonrigid image registration", IEEE Trans. Image Process., vol. 15, no. 3, pp. 529-538, 2006

- [28] Jayaraman, U., Gupta, A.K. and Gupta, P., 2014. An efficient minutiae based geometric hashing for fingerprint database. *Neurocomputing*, 137, pp.115-126.
- 575 [29] Jing, P., Ji, Z., Yu, Y. and Zhang, Z., 2016. Visual search reranking with RElevant Local Discriminant Analysis. *Neurocomputing*, 173, pp.172-180.
- [30] Li, Z., Liu, J., Yang, Y., Zhou, X. and Lu, H., 2014. Clustering-guided sparse structural learning for unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 26(9), pp.2138-2150.