Instance Search with Weak Geometric Correlation Consistency

Zhenxing Zhang, Rami Albatal, Cathal Gurrin, and Alan F. Smeaton

Insight Centre for Data Analytics
School of Computing, Dublin City University
Glasnevin, Co. Dublin, Ireland
{zzhang,cgurrin,asmeaton}@computing.dcu.ie
{rami.albatal}@dcu.ie
https://www.insight-centre.org

Abstract. Finding object instances from large image collections is a challenging problem with many practical applications. Recent methods inspired by text retrieval achieved good results; however a re-ranking stage based on spatial verification is still required to boost performance. To improve the effectiveness of such instance retrieval systems while avoiding the computational complexity of a re-ranking stage, we explored the geometric correlations among local features and 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 and can be applied to all candidate images at a low computational cost. Experimental results on three standard evaluation benchmarks show that the proposed approach results in a substantial performance improvement compared with recent proposed methods.

Keywords: Multimedia Indexing, Information retrieval

1 Introduction

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

Instance search is an interesting yet challenging problem and attracts 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) framework firstly introduced by Sivic and Zisserman [3]. This framework successfully made use of the discriminative power of the local feature descriptors (e.g. SIFT [1], SURF [2]) which are generally robust to changes in image condition and are applied to build a statistical representation 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 speed, even 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 introduced to improve retrieval accuracy. Based on the observation that there can be only one local feature correspondence to any given feature from query object, the geometric layout of objects was adopted to verify the spatial consistency between matched local feature points. Generally, the spatial verification algorithms were applied to refine the ranked results by iteratively optimizing the transformation models and fitting them to the initial correspondences to eliminate inconsistent matches. However those techniques such as RANSAC are normally computationally expensive; they can be applied only as a post-processing step to the top ranked images in the initial result set.

In this 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. Recent work of Jégou et al [5] proposed a novel approach to efficiently apply spatial verification and made it suitable for very large datasets. They used the weak geometric constraints, specifically in the rotational and scale spaces, to examine each individual feature match and filter out those inconsistent feature matches at a very low computational cost. Although it improved retrieval performance for instance search, we observed that their approach considered feature matches independently and ignored the geometric correlation between local features, thus performed less effectively when searching more challenging datasets like FlickrLogos-27 [18]. In this 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.

The experimental results on three standard evaluation benchmarks, in Section 5, illustrate that the proposed method is more reliable, and also more tractable for large image collections, which 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 improve retrieval performance with the geometric information.

Sivic and Zisserman [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 the performance. The work in [11] suggested using very high dimensional vocabulary (1 million visual words) during the quantization process. This method has improved the retrieval precision with more discriminative visual words, and also increased the retrieval efficiency with more sparse image representations, especially for

large scale database. Chum and Philbin [13] brought query expansion techniques to 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 the results in order to remove noisy or ambiguous visual words. Recent works in [7], [10], [9] and [8] extended the BoVW approach by encoding 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.

Recently, alternative approaches have been developed to implicitly verify the feature matches with respect to the consistency of their geometric relations, i.e., scaling, orientation, and location, in the Hough transformation space. Avrithis et al. [12] developed a linear algorithm to effectively compute pairwise affinities of correspondences in 4-dimensional transformation space by applying a pyramid matching model constructed from each single feature correspondence. Jégou et al. [5], increased the reliability of feature matches against imagining condition changes by applying weak constraints to verify the scaling and orientation relations consistency according to the dominant transformation found in the transformation space. Similarly, Zhang et al. [15] proposed to represent the feature points geometric information using topology-base graphs and verified the spatial consistency by performing a graph matching.

Our proposed method follows the direction of implicitly verifying the feature matches 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 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 BoVW architecture, the local features are firstly extracted from each image to encode the invariant visual information into feature vectors. Generally, a feature vector is defined as $\boldsymbol{v}(x, y, \theta, \sigma, q)$, 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 initial matching features $C_{initial}$ could be established by examining feature vector q. The task of spatial verification is to eliminate the unreliable feature matches and only retain the matches set C_{stable} that linked the patches of the same object. The following equation formatted this process:

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

where m_i stands for the ith feature match in the initial match set. f_{sp} stands for the spatial verification function for assessing its geometric consistency. Take

the weak geometric consistency [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 individual feature match and D_θ and D_σ is the dominated transformation in orientation and scale space.

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 follow the same spatial transformation between objects. So instead of verifying the geometric consistency for each match individually, we proposed 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 then define the proposed spatial verification function as following:

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

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

We named our proposed approach to be Weak Geometric Correlation Consistency (WGCC) and Figure 1 demonstrates our idea of using geometric correlations to assess reliability of feature matches. The object of interest (front cover of a box) is highlighted with 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 is consistent between match(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 individually.

3.2 Implementation

To explicitly examine all the correlations between the initial feature matches is a non-trivial problem. If we take a total number of N initial matches as example, the potential pairwise correlation could be modeled as $O(C_N^2)$. The initial feature matching number N is usually large in practical systems, and this will cause a high computational cost to verify all the correlations and makes this solution less attractive for large image collections.
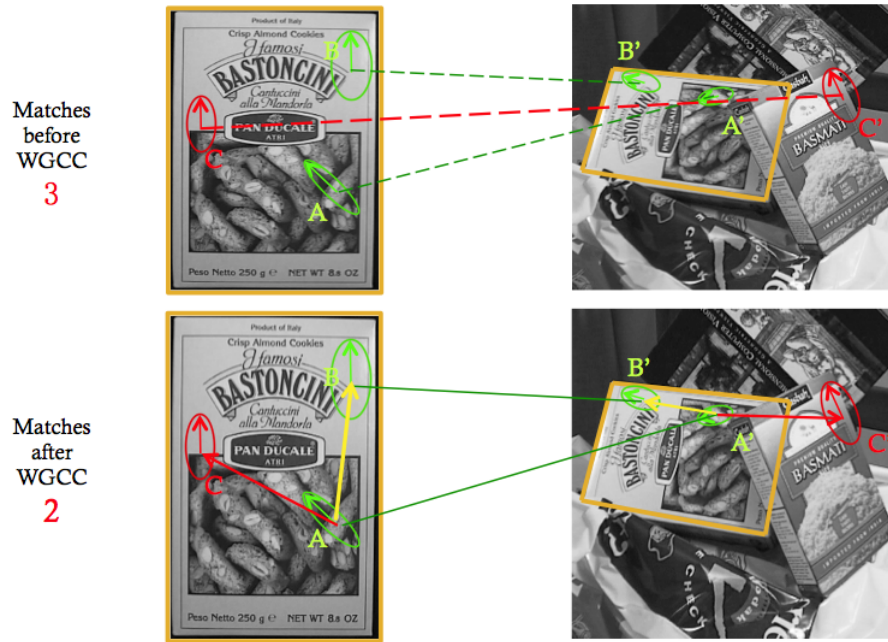
Fig. 1: An illustration of verifying consistency of feature matches using geometric correlations. The green(red) line indicates the consistent(inconsistent) feature matches.

In this work, we proposed a three-step scheme to reduce the complexity of verifying the geometric correlation consistency, and to make it applicable at low cost for large-scale instance search systems. 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:

$$\Delta\theta = \theta_m - \theta_i, \ \Delta\sigma = \sigma_m/\sigma_i \tag{4}$$

In order to reduce the sensitivity to non-rigid deformation, 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 the bin quantization error, each feature match votes to the closest two bins in each parameter space. The Hough voting scheme was applied in searching of the
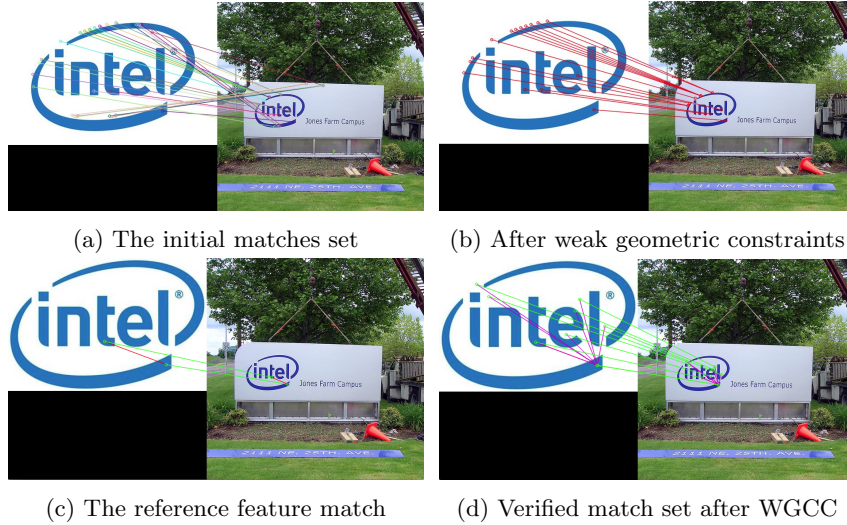
(a) The initial matches set

(b) After weak geometric constraints

(c) The reference feature match

(d) Verified match set after WGCC

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

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 who are 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 step of verifying geometric correlations. We follow the approach of Zhang [15] and adopt a topology-based graph match for this purpose. To represent the topology structure for objects, we created Delaunay Triangulation mesh from the geometric layout among the feature points in object plane. Then we could find the strongest feature matches which corresponding to the the common edges between topology graphs by performing a graph matching.

Verifying weak consistency for geometric correlations. The last step focused on identify the 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 vector $v_{l->n} = (x_l, y_l) - (x_n, y_n)$ 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 vector $v'_{l->n} = (x'_l, y'_l) - (x'_n, y'_n)$. Then the transformation parameters in orientation $\Delta\theta_{i->n}$ and the scale $\Delta\sigma_{l->n}$ between geometric correlations can

be defined as:

$$\Delta\theta_{i->n} = \arccos\frac{\|v_{l->n}\|\|v'_{l->n}\|}{v_{l->n} \cdot v'_{l->n}}, \qquad \Delta\sigma_{i->n} = \frac{\|v'_{l->n}\|}{\|v_{l->n}\|} \qquad (5)$$

It is now possible to assess the spatial consistency by verifying the transformation parameters values with the weak constraints according to equation 3 and further filter out the inconsistent matches to obtain the final set of reliable feature matches.

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. The good news is that we already build weak geometric constraints in the first step to verify the initial feature matches, so only a subset of smaller set 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.

4 Experiments

The goal of experiments is to assess the performance of the proposed weak geometric correlation consistency methods in instance search tasks. Therefore a complete instance search system was developed and comparative experiments were designed to evaluate retrieval performance against state-of-the-art approaches on three standard and publicly available benchmark datasets. The datasets chosen were the Oxford, Pairs6K and FlickrLogis-32 datasets. Each of these datasets includes a set of queries and relevance judgements.

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 the three state-of-the-art approaches.

4.1 Datasets

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

Pairs6K This collection [16] consists of 6,412 images collected by searching for particular Paris landmarks from Flickr. In total, 11 Landmarks with 55 images queries was provided with manually annotated ground truth for users to

evaluate the retrieval performance. The images are considered to be positive if more than 25% of the instance is clearly visible.

FlickrLogos-27 This dataset [18] consists of 5,107 images including 810 annotated positive images corresponding to 27 classes of commercial brand logos and 4,207 distraction images that depict its own logo class. It 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.

4.2 Evaluation protocol

The 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. Subsequently, a visual vocabulary was generated using the approximate K-means algorithm [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 visual words are retrieved from the database collections. Auxiliary information is used to perform the spatial verification to improve retrieval performance.

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

4.3 Experiment Settings

The Weak Geometric Correlation Consistency (WGCC) was compared against the standard BoVW approach as the baseline, but it was also compared against two other advanced approaches; Weak Geometric Consistency(WGC) and Delaunay Triangulation(DT), both of which also adopt the geometric information to enhance the baseline method.

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

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 Hough transformation space.

DT [15]. This approach make use of relations between matched points in a 2-dimensional translation space to improve the matching reliability between

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

Methods	Oxford		Pairs6K		FlickrLogos	
	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

Time[1] measures the average response time per query in second, excluding feature extraction.

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 an point-to-point matches and 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 space and use the product-quantization based algorithm [6] to build up search components for initial feature matching. Then we applied the proposed weak geometric correlation consistency(WGCC) for spatial verification.

5   Results and discussion

Table 1 presents the experimental results of comparing our proposed approach WGCC with the baseline and two enhanced approaches, on the three benchmarks. To study the impact of adopting geometric information for enhancing the retrieval performance in instance search system, we compared the advanced systems against baseline system on the three described datasets. We observed that the advanced approaches for spatial verification consistently improves performance in mAP compared to the baseline. Compared to the other two advanced systems, our proposed approach achieved the best results. Especially on the FlickrLogos dataset, our approach have a 59% relative improvement in the mAP performance from the baseline's 0.145 to 0.231 in our method. This proved that our approach are strong enough to reject inconsistent feature matches, while also flexible for keep the evidence from locally consistent patches.

Figure 3 shows some examples of the improvement in mAP obtained by the proposed WGCC approach compared to the baseline system. The interested object delimited in the yellow box from the query image on left side of each subfigure. The Precision-Recall curve is displayed on the right side with baseline results shown in blue line and WGCC method shown in red. The gap area between two lines indicates the performance improvement for our methods. The high precision value in low recall range indicates that our proposed approaches

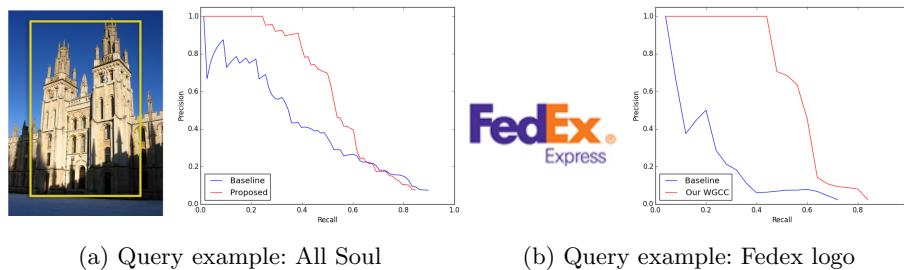(a) Query example: All Soul   (b) Query example: Fedex logo

Fig. 3: Examples of the improvement in the precision-recall curve obtained by using proposed approaches compared to the baseline system.

improved retrieval performance by ranking the good images higher in the ranked list.

Figure 4 displays some good queries results retrieved from the experiment benchmark collections. The results demonstrates the robustness of the proposed methods to the considerable variations in viewpoint, scale, lighting and partial occlusion from practical environment.

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 query task. At run-time, our proposed method achieved comparable retrieval efficiency with the two advanced approaches while providing better accuracy. Although the

Fig. 4: Some good examples of searching experiment benchmark: First row: "Magdalen" in Oxford, Second row: "Ashmolean" in Oxford, Third row: "Ferrari" in FlickrLogo-27, and Fourth row: "Apple" in FlickrLogo-27

retrieval efficiency of WGCC was slightly less than the other approaches, retrieval efficiency could be optimised by adopting parallel computing approaches.

6   Conclusion

This paper proposed novel approach to improve retrieval performance of instance search systems by combining the pairwise geometric correlations with individual feature match transformation to form a weak geometric correlation consistency. This model is strong enough to eliminate inconsistent feature matches while keeping reliable matches using locally spatial correlation. Our experiments shows that our approach consistently outperformed the baseline system in three standard benchmark evaluations and achieved improved results when compared with two advanced systems. This indicates the effectiveness of our method for spatial verification. Another positive aspect is that other advanced technologies, such as automatic query expansion [13], re-ranking base on full spatial verification [11] are compatible with our proposed method and could be used as complementary components to further improve the 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.

7   Acknowledgment

References

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 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