L1-NORM GLOBAL GEOMETRIC CONSISTENCY FOR PARTIAL-DUPLICATE IMAGE RETRIEVAL

Yang Lin, Chen Xu, Li Yang, Zhouchen Lin, Hongbin Zha*

Key Laboratory of Machine Perception (MOE)
School of Electronics Engineering and Computer Science, Peking University, China

ABSTRACT

In all feature point based partial-duplicate image retrieval systems, false matching is a common issue. To tackle the problem, geometric contexts are widely applied to filter the inconsistent matches. This paper presents a novel method called ℓ_1-norm global geometric consistency. We first form the squared distance matrices of all the matched feature points, which remain invariant under translation and rotation between partial-duplicated images. Then we find the scale difference by solving a one-variable ℓ_1-norm error minimization problem, where the large sparse errors correspond to the locations of inconsistent matches. By adopting the Golden Section Search method the minimization problem can be solved efficiently. Extensive experimental results show that our method reaches higher precisions than state-of-the-art geometric verification methods in detecting inconsistent matches. Its speed is also highly competitive even when compared to local geometric consistency based methods.

Index Terms— Image retrieval, geometric verification, partial-duplicate, inconsistent match.

1. INTRODUCTION

In recently years, image search websites, such as Tineye [1], Baidu Shitu [2], and Google Similar Image Search [3], have attracted a lot of researchers and Internet users. Unlike traditional text-based image search, partial-duplicated image search directly use an image rather than words as the retrieval input for searching. Searching images more efficiently and precisely is becoming an essential issue not only for the Internet users but also for many applications, such as image/video copyright violation detection [4], medical diagnosis [5], crime detection [6], and geographical information retrieval [7].

The goal of this work is to improve the accuracy of partial-duplicate image search. Partial-duplicate images are mainly generated by taking photos of the same scene or manually modifying the original pictures using image processing software. Such images usually share almost the same views towards the same object or have slight differences in color tones, lighting conditions, scale changes, rotation transformations, partial occlusions, and complex backgrounds, etc. Since each original image may have many different versions, it is hard to retrieve all of them all from a large scale dataset.

To address the above issue, most of the state-of-the-art partial-duplicate image search methods [8, 9, 10, 11, 12, 13, 14, 15, 16, 17] use local invariant features [18] and Bag-of-Visual-Words representation [9, 10] (a.k.a. Bag-of-Feature, BoF). A typical flow in most of the state-of-the-art methods has two steps: first use SIFT [18] to detect and describe local features in each image, then utilize BoF to index and match features to estimate the similarity between the target image and the retrieved image. Although BoF avoids expensive matching by quantizing the SIFT descriptors into several visual word indices, it also brings some inconsistent feature matches. The inconsistent matches will cause inaccurate similarity between images, which heavily reduces the retrieval performance. To this end, some post-processing methods, such as feature quantization [19], query expansion [20], and geometric verification [11, 12, 13, 14, 16, 17], are used to improve the retrieval accuracy. Among the above, one of the key step is geometric verification, which makes use of geometric prior to verify and filter the inconsistent matches and update the BoF retrieval result.

In this paper, we only focus on the geometric verification step. The main novelties of this paper can be summarized as follows:

- We propose a novel and robust model named ℓ_1-norm global geometric consistency (L1GGC) for detecting inconsistent matches. The partial-duplicate images are allowed to have complex backgrounds and partial-occlusion. They can also have different scale, rotation, and translation from the query image.

- Our model is simple and prior free. We only utilize the coordinates of the feature points to identify the inconsistent matches, whereas most of the existing methods [12, 13, 14, 16, 17] need extra spatial prior for verification, such as characteristic scale and dominate orientation of each SIFT features [18].

*Corresponding Author: Zhouchen Lin, E-mail: zlin@pku.edu.cn.

- Our method is computationally fast. By applying the Golden Section Search to solve our model, our method is more efficient than other global geometric verification approaches [11, 14, 16] and is highly competitive even when compared with existing local geometric verification methods [12, 13, 17].

2. PREVIOUS WORK

In this section, we introduce some state-of-the-art geometric verification methods. There are two kinds of methods for geometric verification. One considers the local geometric consistency of each image. The other tries to model the global geometric consistency.

2.1. Local Geometric Consistency

Most of the local geometric verification methods are based on a similarity transformation model shown in Eq. (1).

$$\begin{bmatrix} x_{2i} \\ y_{2i} \end{bmatrix} = s \cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} x_{1i} \\ y_{1i} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \quad (1)$$

In Eq. (1), (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}) represent the coordinates of the i^{th} corresponding feature points in two images. s, θ, and (t_x,t_y) are the scaling factor, rotation angle, and translation vectors, respectively. To estimate s and θ, Jegou et al. [12] use following equations: $s = \hat{s}_2/\hat{s}_1$, $\theta = \hat{\theta}_2 - \hat{\theta}_1$, where $\hat{s}_1$, $\hat{s}_2$, $\hat{\theta}_1$, and $\hat{\theta}_2$ are the characteristic scale and dominant orientation of two SIFT features, respectively. By heavily relying on the local features properties (e.g., s and θ), some local geometric verification methods are proposed as follows.

Weak Geometric Consistency (WGC) Jegou et al. [12] utilize the peak values of the histograms of s and θ for filtering inconsistent matches. WGC assumes that consistent matches should have similar scale and rotation changes.

Enhanced Weak Geometric Consistency (EWGC) Unlike [12], Zhao et al. [13] use the peak value of the histogram of the translation magnitude to detect inconsistent matches.

Strong Geometric Consistency (SGC) Motivated by WGC [12] and EWGC [13], Wang et al. [17] proposes SGC, which first divides the matches into some groups by rotation changes and utilizes the peak value of the histogram of 2D translation vector as the consistency measurement.

As described in [16], since local geometric verification approaches only verify the consistency within local areas rather than the entire images, they cannot handle geometric inconsistency beyond local areas. Therefore, global geometric verification methods are needed.

2.2. Global Geometric Consistency

Unlike the local geometric verification, the global geometric verification check geometric consistency among all matched features, which can detect inconsistent matches distributed in different local areas. Some state-of-the-art global geometric verification methods are described as follows.

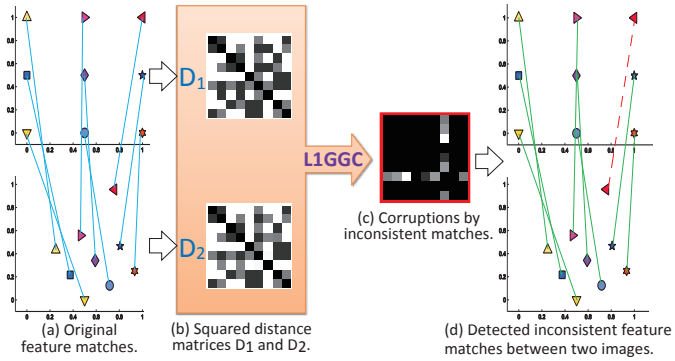

Fig. 1. Our L1GGC method for detecting inconsistent matches. Given initial matches (a), we first compute the matrices that record the squared distances between all matched feature points in each image (b). Then we find the optimal scale that matches the two matrices. The sparse errors in the resulted error matrix (c) reveal the inconsistent matches (dashed line in (d)).

RANSAC As a typical method of global geometric verification, the RANSAC method [11] tries to model the homography transformation among all feature matches between two images. It repeatedly picks a random subset of the whole matches to estimate a transformation, which is time-consuming and hence hard to apply to large scale image retrieval.

Spatial Coding (SC) To break the bottleneck of computation, Zhou et al. [14] recommend to use a spatial coding map to characterize the global spatial relationships among all matches, which is more efficient than RANSAC.

Geometric Coding (GC) Geometric coding [16] is an upgraded version of SC [14]. It performs more efficiently and effectively than SC. The geometric coding map could encode rotation changes effectively and the matches which cause the largest inconsistency will be iteratively removed.

3. OUR APPROACH

In this section, we propose a global geometric verification method called **L1GGC** (ℓ_1 global geometric consistency). We first introduce the basic model of global geometric consistency and then model the violation of global consistency caused by inconsistent matches.

3.1. Modeling Global Geometric Consistency

Suppose we have matched feature points between two images as shown in Figure 1(a). Let $a_{1i} = (x_{1i}, y_{1i})^T$ denote points in the first image and $a_{2i} = (x_{2i}, y_{2i})^T$ denote the matched ones in the second, where $i = 1, \cdots, n$. Then we compute squared distance matrices (See Figure 1(b)):

$$D_1 = \left(\|a_{1i} - a_{1j}\|^2\right)_{i,j=1}^{n}, \quad D_2 = \left(\|a_{2i} - a_{2j}\|^2\right)_{i,j=1}^{n}.$$

In practice, we can compute D_1 and D_2 in a fast way. Take D_1 for example, it can be rewritten as

$$D_1 = \alpha_1 e^T - 2A_1^T A_1 + e\alpha_1^T, \quad (2)$$

where $\alpha_1 = \left(\|a_{1i}\|^2\right)_{i=1}^n$, e is an all-one vector, and $A_1 = [a_{11}\, a_{12}\, \cdots\, a_{1n}] \in \mathbb{R}^{2\times n}$. By this adoption, we can reduce the computing cost from $O(n^2)$ to $O(n)$.

Note that the squared distance matrices have both rotation and translation invariance. So the effect of rotation and translation are naturally removed. If $\{a_{2i}\}$ corresponds to $\{a_{1i}\}$ under a similarity transformation (see Eq. (1)), then we only have to handle the scaling changes between the two images:

$$D_1 = \lambda D_2, \quad (3)$$

where $\lambda > 0$ is a scaling factor.

3.2. Modeling Inconsistent Matches with Sparse Error

When there are inconsistent matches between feature points, Eq. (3) no longer holds. However, suppose the percentage of inconsistent matches is not too high, the discrepancy between D_1 and λD_2 should be sparse, which only resides in the distances from mismatched points to matched ones and those among mismatched ones. Thus we can formulate the following optimization problem:

$$\min_{\lambda>0} \|D_1 - \lambda D_2\|_0, \quad (4)$$

where $\|\cdot\|_0$ represents the ℓ_0-norm (number of nonzero entries in a matrix).

However, (4) is sensitive to noise. A more robust way is to relax ℓ_0-norm to its convex surrogate, ℓ_1-norm, instead:

$$\min_{\lambda>0} \|D_1 - \lambda D_2\|_1, \quad (5)$$

where $\|\cdot\|_1$ represents the ℓ_1-norm (sum of the absolute values of all entries in a matrix).

3.3. Solving by Golden Section Search

An intuitive method to solve (5) is linear programming. However, the computational cost of linear programming depends on the number of unknown variables and linear constraints. In practice, the complexity is known to be quasi-linear with respect to the number of constraints [21]. When the number of feature points increases, the method will be very slow.

Actually, as (5) is a one-variable nonsmooth convex optimization problem, we may use a fast and simple alternative, Golden Section Search [22], to solve it. It has a complexity of $O(log(n))$. Define $\lambda_{ij} = D_{1,ij}/D_{2,ij}$. We first observe that $f(\lambda) = \|D_1 - \lambda D_2\|_1$ is a piecewise linear function, with its corner points being at λ_{ij}. The initial search interval is $[a_0, b_0]$ with $a_0 = \min_{i,j} \lambda_{ij}$ and $b_0 = \max_{i,j} \lambda_{ij}$. The golden search terminates when the search interval contains only one λ_{ij} and this λ_{ij} must be the optimal solution λ^*. Since evaluating $f(\lambda)$ is extremely fast, the computational cost of finding λ^* is very small.

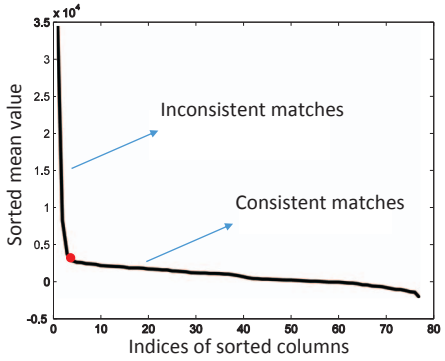

Fig. 2. Detection of inconsistent matches. The big dot indicates the turning point of the curve of mean values of the columns of the error matrix, sorted in a descending order. Large mean values indicate inconsistent matches, while small values correspond to consistent matches.

Once the optimum λ^* is obtained, we can compute an absolute error matrix:

$$E^* = |D_1 - \lambda^* D_2|, \text{ i.e., } E^*_{ij} = |D_{1,ij} - \lambda^* D_{2,ij}|. \quad (6)$$

If there exist inconsistent matches between two images, the entries in the corresponding rows and columns of E^* are supposed to be large (See Figure 1(c)). So we can detect the inconsistent matches by examining the magnitudes in the rows and columns of E^*. Since E^* is symmetric, we may only focus on its columns. We first compute the average value of each column of E^* and sort them in a descending order. We find that there always exists a turning point on the curve (See Figure 2). We take the value of the turning point as threshold, which can be easily detected by finding the peak of the second-difference of the curve. Those columns with values above the threshold correspond to inconsistent matches and the remaining columns indicate consistent matches. In this way, we can detect inconsistent matches effectively (See Figure 1(d)). Then we re-rank all the images by correct matches, using the method by Nister et al. [10].

4. EXPERIMENTS

In this section, in order to prove the effectiveness and efficiency of our approach, we compare our L1GGC method on two benchmark datasets with existing state-of-the-art methods, including the baseline method: BoF [10], two global verification methods: RANSAC [11] and GC [16]; and three local verification methods: WGC [12], EWGC [13], and SGC [17]. We perform the experiments on a server with a 2GHz CPU and 64GB RAM.

4.1. Datasets

We adopt two popular benchmark datasets for evaluation. One is the Holiday dataset and the other is the DupImage

dataset. To make the experiment more realistic and challenging, we also use the MIRflickr1M dataset as distracters.

Holiday dataset The Holiday dataset [12] are mainly personal photos taken on many different scenes (natural or man-made). It has 1491 near-duplicated images in 500 groups.

DupImage dataset Also called GCDup dataset [15]. It has 1104 partial-duplicate images in 33 groups, which are collected from the Internet. Most of the images are composite.

MIRFlickr1M dataset MIRFlickr1M [23] contains one million unrelated images downloaded from Flickr. The image retrieval community often utilizes MIRFlickr to examine a method's scalability and robustness performance by adding different numbers of its images to the benchmark datasets.

The first image of each group in the two benchmark datasets is used as the query image, and the remaining ones in the same group as the expected retrieval result.

4.2. Experiment Settings

We utilize the SIFT method [18] as the key points detector and descriptor. After extracting SIFT descriptors, we train a 100K visual words codebook on the benchmark datasets by using the vocabulary tree method [10] to get the initial feature matches. With the trained codebook, we quantize each 128-dimension feature descriptor into a visual word index. Features that have the same index are considered matched.

4.3. Evaluation Metrics

We use mean average precision (mAP) [11] and average time cost to evaluate the accuracy and speed of various approaches. The mAP is the mean value of each query's average precision: $mAP = \sum_{q=1}^{N_Q} AP(q)/N_Q$. Here N_Q is the number of queries and the measure AP is the area under a precision-recall curve of the image retrieval method mentioned in [11]. It can be computed as follows: $AP = \sum_{i=1}^{N_R}(P(i) \times r(i))/N_A$, where N_R is the number of relevant images, N_A is the number of all images, and $r(i)$ is an indicator function:

$$r(i) = \begin{cases} 1, & \text{the } i^{th} \text{ best match is a relevant image;} \\ 0, & \text{otherwise.} \end{cases}$$

$P(i) = \sum_{j=1}^{i} r(j)/i$ is the precision at the cut-off ranking i in the relevant image list.

4.4. Performance and Discussion

Figure 3 summarizes the mAPs of all methods on the two datasets. As one can see, our L1GGC method outperforms all the methods in comparison. Since BoF does not apply any geometric verifications, it gets the worst results in mAP. We also find that the global geometric methods (RANSAC and GC) perform better than the local ones (WGC, EWGC, and SGC). From the above results, we verify the effectiveness of our global geometric verification method on detecting inconsistent matches.

Table 1 shows the average time costs per image query of all methods on the two datasets. Note that we do not count the time cost of feature extraction, codebook generation, and feature matching, because they are common steps in all the methods. One can see that our method is faster than the other two global verification methods (RANSAC and GC). Especially, the advantage of L1GGC over RANSAC is dramatic. Besides, our method is competitively even when compared with local geometric verification methods (WGC, EWGC, and SGC).

According to the above results, we conclude that our L1GGC is both effective and efficient on detecting inconsistent matches.

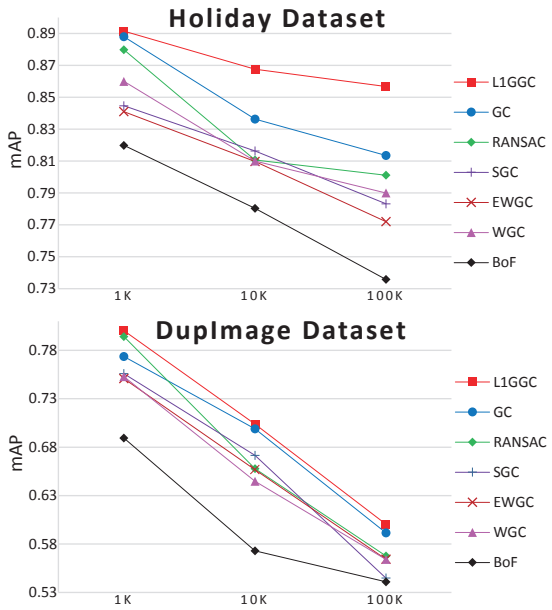

Fig. 3. The mAPs of the compared methods and our L1GGC on two datasets with different numbers of distracters.

Table 1. The average time costs of the compared methods and our L1GGC approach on the two datasets.

	Holiday Dataset	DupImage Dataset
L1GGC	1.03	0.94
RANSAC	18.36	53.64
GC	9.36	5.37
WGC	0.50	0.45
EWGC	0.94	0.41
SGC	1.55	2.23

5. CONCLUSIONS

We have proposed a novel global geometric verification method named L1GGC for partial-duplicate image search. L1GGC is simple, robust, and fast. Based on the ℓ_1-norm Golden Section Search algorithm, L1GGC only utilizes the coordinates of feature points to effectively filter inconsistent matches. In our experiments, L1GGC outperforms state-of-the-art methods with gap on two benchmark datasets with large scale distract images.

Acknowledgements

Z. Lin is supported by NSF China (Grant Nos. 61272341, 61231002, 61121002). H. Zha is supported by the National Basic Research Program of China (973 Program) 2011CB302202.

6. REFERENCES

[1] Idée Inc., "Tineye reverse image search," 2008.

[2] Baidu Inc. Inc., "Baidu shitu image search," 2013.

[3] Google Inc., "Google similar image search," 2009.

[4] M. Douze, H. Jegou, and C. Schmid, "An image-based approach to video copy detection with spatio-temporal post-filtering," *IEEE Transactions on Multimedia*, vol. 12, no. 4, pp. 257–266, June 2010.

[5] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S.C.H. Hoi, and M. Satyanarayanan, "A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 30–44, 2010.

[6] A.J. Lipton, "Keynote: intelligent video as a force multiplier for crime detection and prevention," in *The IEE International Symposium on Imaging for Crime Detection and Prevention*, 2005, pp. 151–156.

[7] W. Li, W. Wang, and F. Lu, "Research on remote sensing image retrieval based on geographical and semantic features," in *International Conference on Image Analysis and Signal Processing*, 2009, pp. 162–165.

[8] J. Tang, H. Li, G.-J. Qi, and T.-S Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," *IEEE Transactions on Multimedia*, vol. 12, no. 2, pp. 131–141, 2010.

[9] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision*, 2003, pp. 1470–1477 vol.2.

[10] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 2161–2168.

[11] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[12] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *European Conference on Computer Vision*, 2008, vol. 5302, pp. 304–317.

[13] W. Zhao, X. Wu, and C. Ngo, "On the annotation of web videos by efficient near-duplicate search," *IEEE Transactions on Multimedia*, vol. 12, no. 5, pp. 448–461, 2010.

[14] W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, "Spatial coding for large scale partial-duplicate web image search," in *Proceedings of the ACM International Conference on Multimedia*, 2010, pp. 511–520.

[15] W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, "Large scale image search with geometric coding," in *Proceedings of the ACM International Conference on Multimedia*, 2011, pp. 1349–1352.

[16] W. Zhou, H. Li, Y. Lu, and Q. Tian, "SIFT match verification by geometric coding for large-scale partial-duplicate web image search," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 9, no. 1, pp. 4:1–4:18, 2013.

[17] J. Wang, J. Tang, and Y. Jiang, "Strong geometrical consistency in large scale partial-duplicate image search," in *Proceedings of the ACM International Conference on Multimedia*, 2013, pp. 633–636.

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

[19] W. Zhou, Y. Lu, H. Li, and Q. Tian, "Scalar quantization for large scale image search," in *Proceedings of the ACM International Conference on Multimedia*, 2012, pp. 169–178.

[20] Y. Kuo, K. Chen, C. Chiang, and W. Hsu, "Query expansion for hash-based image object retrieval," in *Proceedings of the ACM International Conference on Multimedia*, 2009, pp. 65–74.

[21] Q. Ke and T. Kanade, "Robust l_1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 739–746.

[22] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, "Numerical recipes 3rd edition: The art of scientific computing," pp. 492–496, 2007.

[23] M. Huiskes, B. Thomee, and S. Michael, "New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative," in *Proceedings of the ACM International Conference on Multimedia Information Retrieval*, 2010, pp. 527–536.