Low Rank Global Geometric Consistency for Partial-Duplicate Image Search

Li Yang Yang Lin Zhouchen Lin(⊠) Hongbin Zha Key Laboratory of Machine Perception (MOE) School of EECS, Peking University Beijing, P.R. China 100871

{yangli,linyang}@cis.pku.edu.cn, zlin@pku.edu.cn, zha@cis.pku.edu.cn

Abstract-All existing feature point based partial-duplicate image retrieval systems are confronted with the false feature point matching problem. To resolve this issue, geometric contexts are widely used to verify the geometric consistency in order to remove false matches. However, most of the existing methods focus on local geometric contexts rather than global. Seeking global contexts has attracted a lot of attention in recent years. This paper introduces a novel global geometric consistency, based on the low rankness of squared distance matrices of feature points, to detect false matches. We cast the problem of detecting false matches as a problem of decomposing a squared distance matrix into a low rank matrix, which models the global geometric consistency, and a sparse matrix, which models the mismatched feature points. So we arrive at a model of Robust Principal Component Analysis. Our Low Rank Global Geometric Consistency (LRGGC) is simple yet effective and theoretically sound. Extensive experimental results show that our LRGGC is much more accurate than state of the art geometric verification methods in detecting false matches and is robust to all kinds of similarity transformation (scaling, rotation, and translation) and even slight change in 3D views. Its speed is also highly competitive even compared with local geometric consistency based methods.

I. INTRODUCTION

Partial-duplicate image search is to retrieve from a database images which may contain the same contents as the query image. Figure 1 shows some example images. The retrieved images may be altered versions of the query image in color, contrast, scale, rotation, partial occlusion, etc. They may also be scenes taken from slightly different viewpoints. With the rapid development of the Internet and the explosive increase of digital image/video resources, partial-duplicate image search has wide applications, such as image/video copyright violation detection [1], crime prevention, and automatic image annotation [2][3].

Among existing partial-duplicate image search approaches, most of them first utilize the well known bag-of-visual-words (a.k.a. Bag-of-Features, BoF) method [3][4][5][6][7][8][9][10][11][12] to match similar local features between each pair of images. BoF can help avoid expensive evaluation on pairwise similarity between feature points in two images, thus improves search efficiency. BoF based matching is achieved by quantizing local features to visual words and assuming that feature points indexed by the same visual word are matched. However, as only local features are considered the ambiguity of visual words and the feature quantization error result in many false matches among images, leading to low retrieval precision and recall.



Fig. 1. Examples of partial-duplicate web images.

An effective way to improve the matching accuracy is to consider the spatial contexts among the feature points [6]. There are two classes of methods for removing false matches by incorporating spatial contexts. The first class tries to build global geometric consistency. The second depends on local geometric consistency. The conventional method, random sample consensus (RANSAC) [6][13][14], is a representative of the first class, which captures the geometric relationship among all feature points in entire images. However, since RANSAC randomly select a set of matched feature points between two images many times in order to estimate an optimal transformation (homography), it is very slow. So RANSAC is usually adopted for re-ranking the top candidate images. Due to the heavy cost of computing global geometric consistency, the second class aims at reducing the computation cost. These methods only verify spatial consistency among feature points within local areas instead of entire images [11].

In [7], a simpler similarity transformation shown in Eq. (1) replaces the homography transformation used in RANSAC. Such a replacement may lead to only slightly worse results but much faster speed.

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

In Eq. (1), (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}) represent the coordinates of the i^{th} corresponding feature points between the first and second images. s, θ , and t_x and t_y are the scaling factor, rotation angle, and translations in horizontal and vertical directions, respectively. Based on this simplified model, in recent years many approaches, such as [7][8][12][9][10][11],



Fig. 2. The procedure of our LRGGC for detecting false matches. Given initially matched feature points (e.g., by BoF), we first compute the matrices that record the squared distances between all feature points in each image. Then we stack the two distance matrices into one and apply RPCA to decompose it into a low rank matrix and a sparse matrix. The low rank one reveals to correctly matched points, while the sparse one helps identify falsely matched points.

have been proposed to detect falsely matched pairs to improve retrieval performance.

Many of the previous methods, such as [7][8][12], are local geometric verification approaches. They heavily rely on the local feature properties, like SIFT [15] features, especially 1D dominant orientation $\hat{\theta}$ and 1D characteristic scale \hat{s} . Then the parameters θ and s in Eq. (1) could be estimated by the following equations [7]: $s = \hat{s}_2/\hat{s}_1, \theta = \hat{\theta}_2 - \hat{\theta}_1$, where $\hat{s}_2, \hat{s}_1, \hat{\theta}_2$, and $\hat{\theta}_1$ come from two corresponding feature points. In Weak Geometric Consistency (WGC) [7], two histograms are formed from a collection of s and θ , respectively, for all corresponding pairs. The peaks of histograms are used to filter false locally matched features. Obviously, its validity is based on an assumption that correct local matches should have similar local scale and rotation changes. Under the same assumption, Enhanced WGC (E-WGC) [8] utilizes the histogram of ℓ_2 norms of translation vectors instead of scales and orientations. In [12], motivated by WGC and E-WGC Wang et al. proposed Strong Geometric Consistency (SGC). The difference is that SGC divides the matched features into several groups based on orientation changes and uses the histogram of 2D translation vectors.

As mentioned in [11], since these approaches only verify spatial consistency among features within some local areas instead of entire images, they may fail if there exists global geometric inconsistency among local areas. In other words, the local consistency assumption mentioned above may not be satisfied when feature points distributed in different local areas of images. Therefore, *global geometric verification* methods are needed.

As mentioned before, RANSAC is a representative method for global geometric verification. However, its expensive computation cost prevents its utility in large scale image search. So in practice, it is usually applied for re-ranking top-ranked images. To overcome the computation issue, Spatial Coding (SC) [9] is proposed to check spatial consistency globally by using spatial maps to record the spatial relationship among all matched feature points. It makes the global verification much more efficient. However, spatial coding is sensitive to rotation. In order to address this limitation of SC, Geometric Coding (GC) [11] is proposed. GC utilizes the dominant orientation. Besides, it makes full use of SIFT features to build the geometric coding maps to encode the spatial relationship between local feature points and iteratively remove false matches that cause the largest inconsistency. In this way, GC performs both more efficiently and effectively than SC.

In this paper, we propose a new global geometric verification approach called Low Rank Global Geometric Consistency (LRGGC). Our approach is motivated by Robust Principal Component Analysis (RPCA) [16], which has been proven by theory and various applications to be effective in detecting large and sparse noises. We first prove that the matrix that records the squared distances between all feature points in an image is low rank. Suppose the matches are all correct, if we stack the two squared distance matrices into one, it will also be low rank. If there are mismatches, the stacked matrix will differ from a low rank one by a sparse matrix. In this way, we cast the global geometric verification as an RPCA problem, where the low rank matrix models the globally consistent geometric relationship among feature points in images, and the sparse matrix models false matches. We show the procedure of our LRGGC in Figure 2. Our method has three advantages.

- Firstly, LRGGC is a global geometric verification method. So it will be more robust in removing false matches than local geometric verification methods.
- Second, LRGGC is very simple yet theoretically sound. It only utilizes the coordinates of the initially

matched feature points. By its design, LRGGC can naturally handle all kinds of similarity transformation, including scaling, rotation, and translation. Unlike the existing global geometric verification methods, such as SC and GC, LRGGC does not require additional features, like SIFT, and the corresponding histograms, in order to reduce their sensitivity to some transformation, such as rotation. By the theory of RPCA, which LRGGC depends on, LRGGC could detect all false matches correctly with an overwhelming probability if there are sufficient number of initially matched points.

 Third, LRGGC is computationally fast. Its speed is superior to existing global geometric verification methods, such as RANSAC and GC, and is even highly competitive when compared with existing local geometric verification methods, such as WGC, EWGC, and SGC.

In short, LRGGC combines the robustness of global verification methods and the efficiency of local verification methods, making it extremely attractive in large scale duplicate-image search.

In the following, we first introduce our proposed LRGGC in Section II. We give experimental results in Section III to compare the effectiveness and speed of representative global and local verification approaches, using several benchmark datasets. Finally, we conclude the paper in Section IV.

II. OUR APPROACH

In this section, we formulate the problem of detecting false matches as an RPCA problem. We first introduce the basic model of global geometric consistency, and then model the violation of global consistency caused by false matches.

A. Modeling Global Geometric Consistency with a Low Rank Matrix

Suppose we have correct matchings between two images. Let $\mathbf{a}_{1i} = (x_{1i}, y_{1i})^T$ be the feature points in the first image. Their matched points in the second image are denoted as $\mathbf{a}_{2i} = (x_{2i}, y_{2i})^T$, where $i = 1, \dots, n$. For each image, we compute a squared distance matrix $\mathbf{D}_1, \mathbf{D}_2 \in \mathbb{R}^{n \times n}$:

$$\mathbf{D}_1(i,j) = \|\mathbf{a}_{1i} - \mathbf{a}_{1j}\|^2, \quad \mathbf{D}_2(i,j) = \|\mathbf{a}_{2i} - \mathbf{a}_{2j}\|^2$$

We have the following theorem.

Theorem 1. The ranks of the two matrices must satisfy:

$$\operatorname{rank}(\mathbf{D}_1) \leq 4$$
 and $\operatorname{rank}(\mathbf{D}_2) \leq 4$.

Proof: For D_1 , we first write it as

$$\mathbf{D}_1 = \boldsymbol{\alpha}_1 \mathbf{e}^T - 2\mathbf{A}_1^T \mathbf{A}_1 + \mathbf{e} \boldsymbol{\alpha}_1^T,$$

where $\alpha_1 = (\|\mathbf{a}_{1i}\|^2)_{i=1}^n$, e is an all-one vector, and $\mathbf{A}_1 = [\mathbf{a}_{11} \mathbf{a}_{12} \cdots \mathbf{a}_{1n}] \in \mathbb{R}^{2 \times n}$. By this trick, we can reduce the computing complexity from $O(n^2)$ to O(n). Next, we note that

$$\operatorname{rank}(\boldsymbol{\alpha}_1 \mathbf{e}^T) = \operatorname{rank}(\mathbf{e}\boldsymbol{\alpha}_1^T) = 1.$$

and

$$\operatorname{rank}(2\mathbf{A}_1^T\mathbf{A}_1) = \operatorname{rank}(\mathbf{A}_1) \le 2.$$

So we conclude that

$$\operatorname{rank}(\mathbf{D}_1) \leq \operatorname{rank}(\boldsymbol{\alpha}_1 \mathbf{e}^T) + \operatorname{rank}(2\mathbf{A}_1^T \mathbf{A}_1) + \operatorname{rank}(\mathbf{e}\boldsymbol{\alpha}_1^T) \\ \leq 4.$$

That $rank(\mathbf{D}_2) \leq 4$ can be proved in the same way.

Another key observation is that if $\{a_{2i}\}\$ are the corresponding points of $\{a_{1i}\}\$ under a similarity transformation (see Eq. (1)), then we have

$$\mathbf{D}_1 = \lambda \mathbf{D}_2,\tag{2}$$

where $\lambda > 0$ is the scaling factor. Note that the effects of rotation and translation are all removed after computing the squared distances. So if we construct the follow matrix **D** by stacking **D**₁ and **D**₂:

$$\mathbf{D} = \left[egin{array}{c} \mathbf{D}_1 \ \mathbf{D}_2 \end{array}
ight],$$

then $\mathbf{D} \in \mathbb{R}^{2n imes n}$ and

$$\operatorname{rank}(\mathbf{D}) = \operatorname{rank}(\mathbf{D}_1) \le 4. \tag{3}$$

We see that although D may be a big-sized matrix, its rank is actually very low. If $rank(D) \leq 4$, we will have nearly perfect global geometric consistency among all matched pairs in images.

B. Modeling False Matches with a Sparse Matrix

When there are mismatches between feature points, (2) no longer holds. So we cannot expect that (3) is still true. However, suppose the percentage of false matches is not too high, rank(\mathbf{D}) should not deviate from a rank four matrix too much. The discrepancy should only reside in the distances from mismatched points to correctly matched points or the distances among mismatched points, which should account for a relatively small percentage in all squared distances. So we can decompose \mathbf{D} into two matrices:

$$\mathbf{D} = \mathbf{A} + \mathbf{E},\tag{4}$$

where A corresponds to the correct global geometric consistency, whose rank does not exceed four, and E accommodates the corruptions caused by mismatched feature points, which is a sparse matrix.

Thus, our problem can be formulated as decomposing **D** into the sum of a low rank matrix **A** and a sparse error matrix **E**, leading to the following optimization problem:

$$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \gamma \|\mathbf{E}\|_{0}, \text{ s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E},$$
(5)

where $\|\cdot\|_0$ represents the ℓ_0 -norm (number of nonzero entries in a matrix) and $\gamma > 0$ is a parameter that trades off between the rank of **A** and the sparsity of **E**.

C. Robust Principal Component Analysis

Eq. (5) is exactly the RPCA problem [16]. It is an NP-hard problem [16]. So Candès et al. proposed to solve its convex surrogate instead:

$$\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \gamma \|\mathbf{E}\|_1, \text{ s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E},$$

where $\|\cdot\|_{*}$ and $\|\cdot\|_{1}$ represent the nuclear norm (sum of the singular values of a matrix) and ℓ_1 -norm (sum of the absolute values of all entries in a matrix), which are known to be the convex envelopes of rank and $\|\cdot\|_0$, respectively [16]. This convex program is called Principal Component Pursuit (PCP). Candès et al. proved that under quite general conditions, e.g., the rank of A is not too high and the number of non-zero entries in E is not too large, the ground truth A and E can be recovered with an overwhelming probability (the probability of failure drops to zero exponentially when the matrix size increases) [16][17]. Note that it is remarkable that the magnitudes in E is not assumed. Therefore, our proposed LRGGC method is theoretically guaranteed to detect the false matches no matter how wildly the false matches are, if the number of initially matched points is sufficiently large and the percentage of false matches is not too high.

PCP can be very efficiently solved by the alternating direction method (ADM) [18][19]. In particular, a few iterations are sufficient to obtain highly accurate estimates of false matches.

In Figure 2, after false match detection by LRGGC, we simply count the number of correct matches as the similarity between the query and the retrieved images to rank all the images in a database.

III. EXPERIMENTS

In this section, we compare our LRGGC approach with existing state of the art geometric verification methods on two benchmark datasets, in order to validate the effectiveness and efficiency of our method. All the experiments are performed on a desktop computer with a 3.8 GHz CPU and 8 GB RAM.

A. Experiment Setup

1) Datasets: Two popular datasets, Holiday dataset [7] and DupImage dataset [10], are used for evaluation. We also adopt the MIRflickr1M [20] dataset as distracters.

- *Holiday dataset.* Holiday contains 1491 high resolution near-duplicated images from original personal photographs taken on holidays. The dataset has totally 500 groups of images, which covers a very large variety of scene types (natural, man-made, etc.). The first image of each group is used as the query image, and the rest of the images in the same group should be ranked at the top of the search results.
- *DupImage dataset.* DupImage contains 1104 partialduplicate images collected from the Internet. Most of the images are composite. They are manually divided into 33 groups. In each groups, images are partial duplicates of each other. In the same way, we choose the first image in each groups as the query, and expect that all the top-ranked images in the same group.
- MIRFlickr1M dataset. MIRFlickr1M has one million un-classified images downloaded from Flickr. The image retrieval community often use MIRFlickr to test the robustness and scalability of different approaches by randomly adding images from this dataset to other datasets. This would make image retrieval more challenging and realistic.

2) Local features: We use SIFT [15] as the standard feature. SIFT detects Difference-of-Gaussian (DoG) key points and extracts a 128-dimensional SIFT descriptor for each key point. It also provides the scale and dominant orientation information of the key points. After extracting SIFT descriptors, we adopt a codebook of 200K visual words to generate initially matched local feature pairs by quantizing each SIFT descriptor to a visual words index. The scale, orientation, and coordinates of each SIFT feature are preserved for further processing.

3) Evaluation Metrics: We adopt the widely used mean Average Precision (mAP) [6] and average time cost to evaluate the accuracy and speed of various approaches. The mAP for a set of queries is the mean of the average precision scores for each query. It can be computed as follows:

$$mAP = \sum_{q=1}^{Q} AP(q)/Q$$

Here Q is the number of queries and the measure AP [6] is the area under a precision-recall curve of the image retrieval method. It is computed as follows:

$$AP = \sum_{i=1}^{N} (P(i) \times r(i))/M,$$

where N is the number of relevant images, M is the number of all images, and r(i) is an indicator function. r(i) = 1 if the the *i*-th best match is a relevant image; r(i) = 0 if not. $P(i) = \sum_{j=1}^{i} r(j)/i$ is the precision at the cut-off ranking *i* in the relevant image list.

B. Compared Methods

Several state of the art geometric verification methods are chosen for comparison. They include the standard BoF method [5], RANSAC [6], GC [10], WGC [7], E-WGC [8], and SGC [12]. We use BoF as a basline. RANSAC and GC are global verification methods and WGC, E-WGC, and SGC are local verification methods.

C. Performance and Discussion

Figure 3 shows the mAPs of all the methods. Our proposed LRGGC method significantly outperforms all the state of the art methods. BoF, as a baseline method, performs the worst. The reason is that it does not use any geometric consistency verification. So its results have a lot of false matches. The local verification methods, WGC, E-WGC, and SGC, perform not as well as the global verification methods, RANSAC and GC. This testifies to the validity of using global geometric consistency to detect false matches. Especially, since RANSAC is based on more powerful affine transformations, which is time-consuming, it could only be applied on the top results. So its performance is highly determined by the recall of top results.

Figure 4 summarizes the average time costs for one image query on the two datasets by the compared methods and our LRGGC approach. The time costs of SIFT feature extraction and visual word index generation are not included because they are shared by all methods, which are for detecting false matches only. As one can see, our method is faster than



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



Fig. 4. The average time costs of the compared methods and our LRGGC approach on the two datasets.

other two global verification methods, RANSAC and GC. In particular, the advantage of LRGGC over RANSAC is significant and on the Holiday data set GC's cost is three times as long as our LRGGC. Our method is also highly competitive even compared with local verification methods, WGC, E-WGC, and SGC.

Figure 3 has shown that global methods all perform better than the local ones. On the other hand, Figure 4 shows that RANSAC is too slow for practical applications. Therefore, we only compare with the remaining global method, GC, on image retrieval performance. As shown in Figure 5, two images are selected as queries to demonstrate the performance of image retrieval. Compared with GC, our approach has much higher mAPs (Figure 5(a) and (b)) and more relevant images are ranked to the top (Figure 5(c) and (d)).

By the above experiments, we confirm that our LRGGC is both more robust and fast, making it highly attractive in partial-duplicate image search.

IV. CONCLUSIONS

We have proposed a new global geometric verification method called LRGGC for partial-duplicate image search. LRGGC only uses the coordinates of feature points to efficiently and robustly filter false matches. Our method is simple yet effective and fast. The underlying RPCA model further provides LRGGC a solid theoretical ground. The proposed LRGGC significantly outperforms state of the art methods in our experiments on two benchmark datasets. In the future, we will investigate whether LRGGC can help *correct* false matches. We will also target on further speeding up LRGGC.

ACKNOWLEDGMENT

Z. Lin is supported by NSF China (Grant Nos. 61272341, 61231002, 61121002) and MSRA. H. Zha is supported by the National Basic Research Program of China (973 Program) 2011CB302202. The authors would like to thank Wengang Zhou for providing their DupImage datasets.

REFERENCES

- Google Similar Image Search, http://googleblog.blogspot.com/2009/10/similarimages-graduates-from-google.html, 2009.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, *Image retrieval: ideas, influences, and trends of the new age.* ACM Computing Surveys, 40(2):1-60, 2008.
- [3] J. Tang, H. Li, G.-J. Qi, and T.-S. Chua Image annotation by graph-based inference with integrated multiple/single instance representations. IEEE TMM, 12(2):131-141, 2010.



Fig. 5. Comparison between GC and our LRGGC on the image retrieval performance. (a)-(b) The Precision-Recall curves for the queries "iPhone" and "Singer", respectively. (c)-(d) The top retrieved images for the queries "iPhone" and "singer" (rank one to rank ten), respectively. The false relevant images are marked with red triangles at the bottom.

- [4] J. Sivic and A. Zisserman, Video google: A text retrieval approach to oject matching in videos. ICCV, 2003.
- [5] D. Nister and H. Stewenius, Scalable recongnition with a vocabulary tree. CVPR, 2006.
- [6] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, Object retrieval with large vocabularies and fast spatial matching. CVPR, 2007.
- [7] H. Jegou, M. Douze, and C. Schmid, Hamming embedding and weak geometric consistency for large scale image search. ECCV, 2008.
- [8] W.-L. Zhao, X. Wu, and C.-W. Ngo, On the annotation of web videos by efficient near-duplicate search. IEEE TMM, 12(5):448-461, 2010.
- [9] W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, Spatial coding for large scale partialduplicate web image search. ACM MM, 2010.
- [10] W. Zhou, H. Li, Y. Lu, and Q. Tian, Large scale image search with geometric coding. ACM MM, 2011.
- [11] W. Zhou, H. Li, Y. Lu, and Q. Tian, Sift match verification by geometric coding for large-scale partial-duplicate web image search. ACM TOMCCAP, 9(1), 2013.
- [12] J. Wang, J. Tang, and Y. Jiang, Strong geometry consistency for large scale partialduplicate image search. ACM MM, 2013.
- [13] M. A. Fischler, and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. CACM, 24(6):381-395, 1981.
- [14] O. Chum, J. Matas, and S. Obdrzalek, Enhancing RANSAC by generalized model optimization. ACCV, 2004.

- [15] D. Lowe, Distinctive image features from scale-invariant key points. IJCV, 30(2):77-116, 2004.
- [16] E. Candes, X. Li, Y. Ma, and J. Wright, Robust principal component analysis? JACM, 2011.
- [17] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky, Sparse and lowrank matrix decompositions. IFAC SYSID, 2009.
- [18] Z. Lin, R. Liu, and Z. Su, *Linearized alternating direction method with adaptive penalty for low rank representation.* NIPS, 2011.
- [19] M. Tao, and X. Yuan, Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIOPT, 21(1):57-81, 2011.
- [20] M. J. Huiskes, B. Thomee, and M. Lew, New trends and ideas in visual concept detection. ACM MIR, 2010.