

# Logo Recognition via Improved Topological Constraint

Panpan Tang and Yuxin Peng<sup>(✉)</sup>

Institute of Computer Science and Technology, Peking University,  
Beijing 100871, China  
panpantangbjtu@gmail.com, pengyuxin@pku.edu.cn

**Abstract.** Real-world logo recognition is challenging mainly due to various viewpoints and different lighting conditions. Currently, the most popular approaches are usually based on bag-of-words model due to their good performance. However, their shortcomings lie in two main aspects: (1) wrong recognition results caused by mismatching of keypoints. (2) high computational complexity and extra noise caused by a large number of keypoints which are irrelevant to the target logo. To address these two problems, we propose a new approach which combines feature selection and topological constraint for logo recognition. Firstly, feature selection is applied to filter out most of the irrelevant keypoints. Secondly, an improved topological constraint, which considers the relative position between a keypoint and its neighboring points, is proposed to reduce the number of mismatched keypoints. It is proven in this paper that the proposed constraint can remove the keypoints which are not on the same planar surface with the others from the  $k$  nearest neighbors of a keypoint. This property is very important to logo recognition because logos are planar objects in real world. The proposed approach is evaluated on two challenging logo recognition benchmarks, FlickrLogos-32 and FlickrLogos-27, and the experimental results show its effectiveness compared to other popular methods.

**Keywords:** Logo recognition · Improved topological constraint · Feature selection

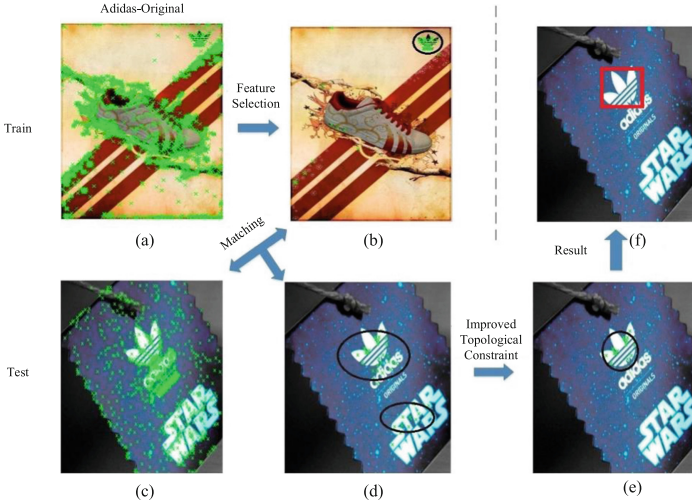
## 1 Introduction

Logo recognition is a sub-problem of object recognition, and has attracted increasing interests in recent years because of its commercial benefits such as measurement of brands' exposure. Given an image, the goal of logo recognition is to check whether it contains any logo and to determine where the logos are located. In real world, affine transformations such as rotation, shearing and scaling caused by variety of viewpoints make logo recognition a challenging task. Besides, occlusion, different lighting conditions as well as diverse appearance of logo itself increase the difficulty of logo recognition.

Recently, a number of works have tackled logo recognition using bag-of-words (BoW) model [1–7]. These methods firstly extract local features such as SIFT or SURF from images. Then these features of keypoints are clustered and quantized into individual integer numbers which are called visual words, and an image is represented as a collection of visual words, known as bag-of-words (BoW). Finally, the similarity between a logo image and a test image is measured based on these visual words. Compared with the original features, the quantized features are more suitable for large scale image retrieval/recognition systems.

However, approaches based on BoW model have two common shortcomings. The first one is that the discriminative ability of visual words decreases in some degree due to quantization, causing that two keypoints belonging to the same visual word may locate at different objects or different regions of the same object, which is called mismatching. To enhance the discriminative ability of visual words and reduce mismatching, Zhou et al. [8] proposed a novel scheme, named spatial coding, to encode the spatial relationships among local features in an image. Since it is specifically for partial-duplicate image search and is based on the assumption that the query image and matched images share the same or very similar spatial layout, it is not inherently suitable for logo recognition in real-world images. Romberg et al. [4] proposed a Bundle min-Hashing (BmH) approach, which aggregates individual local features with the features from their spatial neighborhood into bundles. However, it ignores the relative position between features in a bundle, which limits the discriminative ability in some degree. Kalantidis et al. [1,3] proposed to locally group features in triples and represent the triangles by the signatures capturing both visual appearance and local geometry. However, the geometric constraint in [1,3] is not strictly affine invariant and occasionally fails to detect logos especially with rotation changes which are typically the case for logos in real-world images. Wan et al. [5] proposed a Tree-based Shape Descriptor (TSD) to encode the shape of logos by depicting both the appearance and spatial information of four local key-points. Although the Tree-based Shape Descriptor is strictly invariant to affine transformation in real-world images, it is easily affected by the loss of keypoints which commonly happens in the case of occlusion. In this paper, we propose an Improved Topological Constraint (ITC) which considers the relative position between a keypoint and its neighboring points. Moreover, we apply the cyclic Longest Common Subsequence (LCS) in ITC for similarity measurement in order to be robust to the loss of keypoints.

The second problem is that a large number of keypoints in an image are irrelevant to the target logo, which increase computational complexity as well as bring some noise. An optional solution is by only considering keypoints in the mask region. However, not all keypoints in the mask region is relevant to the target logo. Additionally, the number of keypoints in different mask regions varies a lot, which significantly affects the result of recognition. In our approach, we apply feature selection based on Mutual Information (MI) to filter out most irrelevant keypoints.



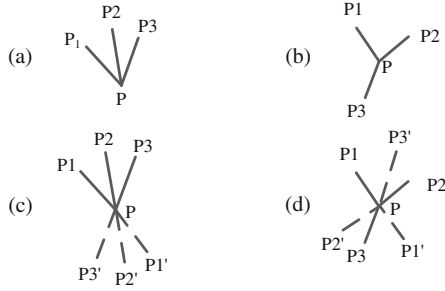
**Fig. 1.** Logo recognition process of our approach (taking Adidas-Original as an example). (a)→(b), most irrelevant keypoints are filtered out by feature selection; (b), (c)→(d), keypoints matching process, note that some mismatched keypoints appears in (d) (keypoints in the smaller ellipse); (d)→(e), mismatching is removed by the proposed improved topological constraint; (e)→(f), Adidas-Original logo is recognized correctly)

The contributions of this paper are twofold: (1) we propose an improved topological constraint to reduce the number of mismatched keypoints. (2) we propose a new approach for logo recognition by combining feature selection and the proposed topological constraint. Figure 1 shows the process of the proposed approach.

## 2 Improved Topological Constraint

Tell et al. [12] proposed a topological constraint which performs string matching to ensure one-to-one matching and to preserve cyclic order. Firstly, they represent the  $k$  nearest neighbors (kNN) of a point as a string according to their cyclic order, as shown in Fig. 2. Then for a pair of matched points, they get the cyclic Longest Common Subsequence (LCS) of their kNN using the algorithm proposed by Gregor et al. [10]. Finally, two points are truly matched if and only if they have longer cyclic LCS than any other pair of points that contain at least one of them. This constraint is invariant to affine transformation if the profile is on a planar surface [12].

The constraint proposed by Tell et al. [12] which combines local and semi-local feature comparison has good discriminative power. But it only considers the cyclic order of kNN and ignores the angular relationship between them. The following is a lemma proven by Wan et al. [5].

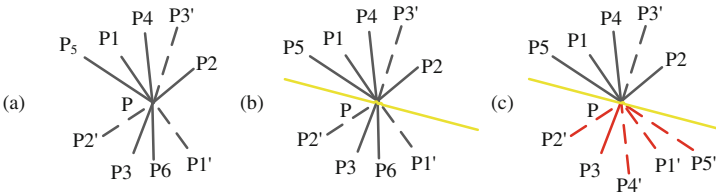


**Fig. 2.** The cyclic order of the points in (a) and (b) are both  $P_1P_2P_3$ . The difference between them can not be reflected in the cyclic order. After adding the symmetric points such as in (c) and (d),  $P_1, P_2, P_3$  will not appear in the cyclic LCS together.

**Lemma 1.** *As shown in Fig. 2(a), if  $PP_2$  lies in  $\angle P_1PP_3$  and  $\angle P_1PP_3$  is less than  $\pi$ ,  $PP_2$  is still in  $\angle P_1PP_3$  after affine transformation.*

As shown in Fig. 2, the cyclic order of the points in Fig. 2(a) and (b) are both  $P_1P_2P_3$ , so their cyclic LCS is  $P_1P_2P_3$  whose length is 3. However, in Fig. 2(b),  $PP_2$  is not in the  $\angle P_1PP_3$  that is less than  $\pi$ , which means that the obtained cyclic LCS does not satisfy Lemma 1, and  $P_1, P_2, P_3$  can not be on the same planar surface.

To address this problem, we propose an Improved Topological Constraint (ITC). As shown in Fig. 2(c) and (d), for each neighboring point of the point  $P$ , we add an extra symmetric point.  $P'_1, P'_2, P'_3$  are the symmetric points of  $P_1, P_2, P_3$  respectively. Then the cycle orders of the points are changed into  $P_1P_2P_3P'_1P'_2P'_3$  and  $P_1P'_3P'_2P'_1P_3P'_2$  respectively. Then their cyclic LCS becomes  $P_1P_2P'_1P'_2$ . Finally, we remove the added points  $P'_1, P'_2$  from the cyclic LCS and it becomes  $P_1P_2$ , whose length is 2. Compared with the original topological constraint, the proposed improved topological constraint removes point  $P_3$  which is not on the same planar surface with the others. Furthermore, it can be proven that there exists at least one cyclic LCS obtained by ITC can satisfy that any three points of it meet Lemma 1.



**Fig. 3.** (a) is the original original cyclic LCS; In (b), the yellow line which crosses  $P$  divide all points into two parts; In (c), points in the under part are replaced by the symmetry points of the other part (Color figure online).

**Theorem 1.** *After adding the symmetric points, there exists at least one cyclic LCS satisfying that all points appear in pairs, e.g., if  $P_i$  appears, its symmetric point  $P'_i$  appears in the cyclic LCS too.*

*Proof.* Suppose that there exists a cyclic LCS of Fig. 3(a), denoted as  $L$ .

1. We can always use a line  $s$  which crosses the central point  $P$  but does not cross any other point to divide  $L$  into two parts. The yellow line in Fig. 3(b) is one of these lines.

2. As point  $P$  is the central point, if point  $P_i$  and its symmetric point  $P'_i$  both appear in  $L$ , they will locate at different part.

3. If we replace all the points in the under part with the symmetric points of the other part, we can get another subsequence  $L'$ , as shown in Fig. 3(c).

4. As  $Length(L') \geq Length(L)$ ,  $L'$  must be a cyclic LCS too. And all points of  $L'$  appear in pairs.

5. As described above, we can always find a cyclic LCS which satisfies that all the points appear in pairs after adding the symmetric points.  $\square$

**Theorem 2.** *If a cyclic LCS satisfies that all the points appear in pairs, any three points of it will satisfy Lemma 1.*

*Proof.* Suppose there are three points of a cyclic LCS which satisfies that all the points appear in pairs, cyclic LCS of these three points and their symmetric points will not exceed four points, as shown in Fig. 2(c) and (d). It signifies that these three points and their symmetric points can not appear in a cyclic LCS together. So the hypothesis is invalid and any three points will satisfy Lemma 1.  $\square$

Through Theorems 1 and 2 we can prove that there exists at least one cyclic LCS obtained by ITC can satisfy that any three points of it meet Lemma 1.

## 3 Logo Recognition

### 3.1 Feature Selection

The first step of our approach for logo recognition is feature selection. It has two advantages: Firstly, it can filter out a lot of keypoints in the test images which are irrelevant to the target logo, thus helps to improve the recognition accuracy. Secondly, it can greatly reduce the number of keypoints which need to be matched in the following step, and thus is helpful to improve the recognition speed.

A common feature selection method based on the expected Mutual Information (MI) of term  $t$  and class  $c$  is adopted in this paper. MI measures how much information the presence/absence of a term contributes to make the correct classification decision on  $c$ . For a term  $t$  and a category  $c$ , their MI is equivalent to Eq. 1 [11].

$$I(t, c) = \frac{N_{11}}{N} \log_2 \frac{N N_{11}}{N_1 N_{.1}} + \frac{N_{01}}{N} \log_2 \frac{N N_{01}}{N_0 N_{.1}} \\ + \frac{N_{10}}{N} \log_2 \frac{N N_{10}}{N_1 N_{.0}} + \frac{N_{00}}{N} \log_2 \frac{N N_{00}}{N_0 N_{.0}} \quad (1)$$

---

**Algorithm 1.** Feature selection for a positive sample containing logo  $c$

---

**Input:**

- $I = \{P_i\} :=$  All BoWs in the positive sample
- $S = \{I_j\} :=$  All images in the training set
- $R = \{I_k\} :=$  All images in the training that contain logo  $c$
- $k :=$  Number of BoWs to be selected

**Output:**

- $Q :=$  The top  $k$  features of  $I$
  - 1:  $V \leftarrow \square$
  - 2: **for each**  $P_i \in I$
  - 3: **do**  $N_{11} \leftarrow 0, N_{10} \leftarrow 0, N_{01} \leftarrow 0, N_{00} \leftarrow 0$
  - 4:     **for each**  $I_j \in S$
  - 5:         **if**  $P_i \in I_j$
  - 6:             **then if**  $I_j \in R$  **then**  $N_{11} \leftarrow N_{11} + 1$
  - 7:             **else**  $N_{10} \leftarrow N_{10} + 1$
  - 8:         **else**
  - 9:             **if**  $I_j \in R$  **then**  $N_{01} \leftarrow N_{01} + 1$
  - 10:            **else**  $N_{00} \leftarrow N_{00} + 1$
  - 11:      $N \leftarrow \#(S)$
  - 12:      $MI \leftarrow$  compute  $I(t, c)$  according to Eq. 1
  - 13:     Append( $V, \langle P_i, MI \rangle$ )
  - 14: Descending order  $V$  by  $MI$
  - 15:  $Q \leftarrow$  first  $k$  elements of  $V$
  - 16: **return**  $Q$
- 

where the  $N_s$  are counts of images that have the values of  $e_t$  and  $e_c$  that are indicated by the two subscripts. For example,  $N_{10}$  is the number of images that contain  $t$  ( $e_t = 1$ ) and are not in  $c$  ( $e_c = 0$ ).  $N_{1.} = N_{10} + N_{11}$  is the number of images that contain  $t$  ( $e_t = 1$ ).  $N = N_{00} + N_{01} + N_{10} + N_{11}$  is the total number of images.

In this paper, category  $c$  is logo type and term  $t$  is visual word. For each logo in the training set, the images containing the logo are positive samples and the others are negative samples. For each positive sample, we compute the MI of each visual word according to Eq. 1. The top  $k$  visual words whose MI are greater than the others are retained finally, since visual words with greater MI have more relevance to the target logo. In practice, the top  $k$  visual words are mainly within the target logo region, as shown in Fig. 1(b), thus most irrelevant keypoints are filtered out. The detailed algorithm for feature selection is shown in Algorithm 1.

### 3.2 Recognition

The second step of our approach is recognition. Denote an image  $I_q$  and a matched image  $I_t$  are found to share  $N$  pairs of matched keypoints<sup>1</sup>. Then the corresponding k-Nearest-Neighbors (kNN) of these matched keypoints for both

---

<sup>1</sup> two keypoints are matched if they are quantized into the same visual word.

$I_q$  and  $I_t$  can be generated and denoted as  $kNN_q$  and  $kNN_t$ . As the Improved Topological Constraint (ITC) described above, we add an extra symmetric point for each point in  $kNN_q$  and  $kNN_t$ , as shown in Fig. 2, and the points of  $kNN_q$  and  $kNN_t$  are doubled, denoted as  $kNN'_q$  and  $kNN'_t$ . Then  $kNN'_q$  and  $kNN'_t$  are represented as strings according to cyclic order and their cyclic Longest Common Subsequence (LCS) is computed using the algorithm proposed by Gregor et al. [10], denote as  $LCS(kNN'_q, kNN'_t)$ . Ideally, if all  $N$  matched pairs are true, the length of  $LCS(kNN'_q, kNN'_t)$  is equivalent to the size of  $kNN'_q$  or  $kNN'_t$ , but if some false matches exist, the former will be smaller than the latter. The similarity between two matched keypoints is defined as Eq. 2:

$$r = \frac{\text{Length of } LCS(kNN'_q, kNN'_t)}{\min\{\#(kNN'_q), \#(kNN'_t)\}} \quad (2)$$

Two keypoints are truly matched if  $r$  is bigger than a predefined threshold  $\alpha$ , which controls the strictness of topological constraint and impacts the verification performance.

We formulate the logo recognition as a voting problem. Each matched keypoint in the test image votes on its matched image. Intuitively, the MI weight of feature selection can be used to distinguish different matched keypoints. However, from our experiments, we find that simply counting the number of matched keypoints which are quantized to different visual words yields similar or better results. To recognize if the test image contains any logo and which logo does it contain, it need to be matched with every image in the training set, which are regarded as reference images. Since those reference images have been pre-processed by feature selection in the last step and only a few irrelevant keypoints are kept, the matching process is quite fast. Denote test image is  $I_q$ , reference images are  $S = \{I_i\}$  and corresponding numbers of matched features are  $M = \{m_i\}$ . Then the recognition result is define as Eq. 3:

$$c_q = \begin{cases} \{c_t \mid m_t = \max(M)\}, & \text{if } \max(M) \geq \beta \\ \text{no-logo}, & \text{if } \max(M) < \beta \end{cases} \quad (3)$$

where  $c_q$  is the recognition result of  $I_q$ ,  $c_t$  is the logo class of reference image  $I_t$ , and  $\beta$  is a predefined threshold which determines whether an image contains any logo.

The detailed algorithm for recognition with improved topological constraint is shown in Algorithm 2.

## 4 Experiments

We evaluate the proposed approach on two challenging and commonly used logo recognition datasets, FlickrLogos-32 [3] and FlickrLogos-27 [1]. They are specially designed for real-world logo recognition. FlickrLogos-32 contains photos showing brand logos and is used for the evaluation of logo retrieval and multi-class logo detection/recognition systems on real-world images. It contains 8240

**Algorithm 2.** Recognition for test image  $I_q$ **Input:**

$I_q = \{P_i\}$  := BoW features of  $I_q$   
 $S' = \{I_j\}$  := Images pre-processed by feature selection in the training set  
 $C = \{c_j\}$  := Corresponding logo class of images in  $S'$   
 $\alpha$  := Similarity threshold of matched keypoints  
 $\beta$  := Similarity threshold of images

**Output:**

$c_q$  := The recognition result of  $I_q$   
 1:  $M \leftarrow \square$   
 2: **for each**  $I_j \in S'$   
 3: **do**  $V \leftarrow \square$   
 4:     **for each**  $P_i \in I_q$   
 5:     **do for each**  $P_k \in I_j$   
 6:         **do if**  $P_i == P_k$  **then** Append( $V, \langle P_i, P_k \rangle$ )  
 7:      $T \leftarrow \square$   
 8:     **for each**  $\langle P_i, P_k \rangle \in V$   
 9:     **do**  $kNN_i \leftarrow$  k-Nearest-Neighbors of  $P_i$   
 10:          $kNN_k \leftarrow$  k-Nearest-Neighbors of  $P_k$   
 11:          $kNN'_i \leftarrow$  adding extra symmetry points to  $kNN_i$  as Fig. 2  
 12:          $kNN'_k \leftarrow$  adding extra symmetry points to  $kNN_k$  as Fig. 2  
 13:          $s_i \leftarrow$  cyclic order of  $kNN'_i$   
 14:          $s_k \leftarrow$  cyclic order of  $kNN'_k$   
 15:          $l \leftarrow$  cyclic Longest Common Subsequence of  $s_i$  and  $s_k$   
 16:          $r \leftarrow \text{Length}(lcs) / \min\{\#(kNN'_i), \#(kNN'_k)\}$   
 17:         **if**  $r \geq \alpha$  **then** Append( $T, P_i$ )  
 18:      $m_j \leftarrow$  number of distinct visual words in  $T$   
 19:     Append( $M, m_j$ )  
 20:  $c_q \leftarrow$  no-logo  
 21: **if**  $\max(M) \geq \beta$   
 22: **then**  $c_q \leftarrow \{c_t | m_t = \max(M)\}$   
 23: **return**  $c_q$

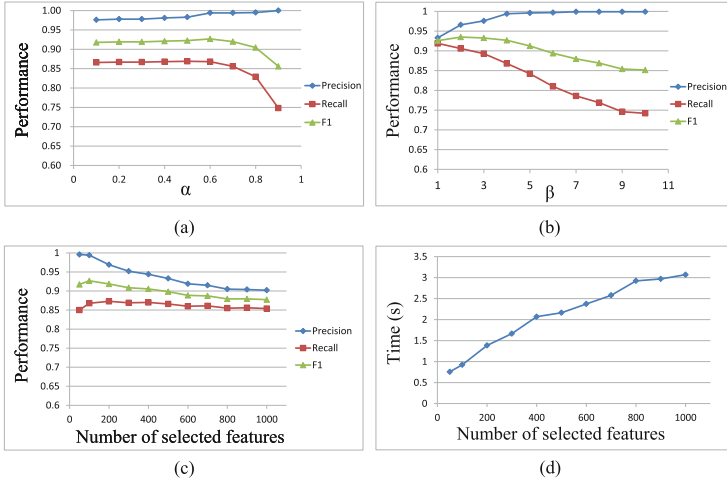
images and is split into three disjoint subsets, each containing images of 32 logo classes. FlickrLogos-27 dataset is an annotated logo dataset downloaded from Flickr. Different from FlickrLogos-32, it contains more than four thousand classes in total.

#### 4.1 Impact of Parameters

The performance of the proposed approach is affected by three main parameters: the similarity threshold of matched keypoints  $\alpha$ , the similarity threshold of images  $\beta$  and the number of selected features. We evaluate the impact of these parameters on FlickrLogos-32 dataset, since the large number of images and distractors makes this dataset more analogous to natural scenario.

The similarity threshold of matched keypoints  $\alpha$  controls the strictness of the topological constraint. The performance of precision, recall and F1 score for





**Fig. 4.** (a) Performance with different similarity threshold of matched keypoints  $\alpha$ ; (b) Performance with different similarity threshold of images  $\beta$ ; (c) Performance with different number of selected features; (d) Time cost for the recognition of an image with different number of selected features.

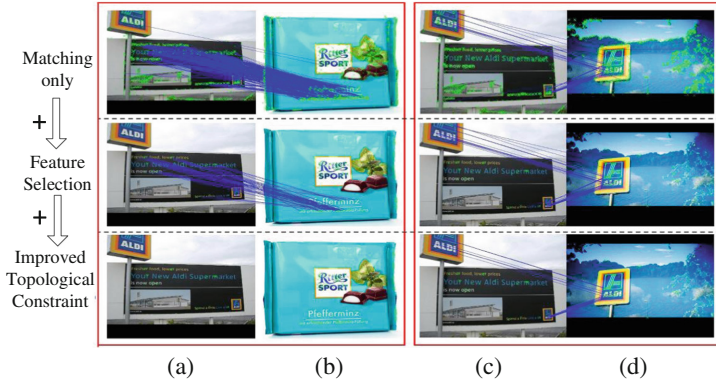
different values of  $\alpha$  is shown in Fig. 4(a) (with  $\beta = 4$  and the number of selected features equal to 100). Note that  $\alpha = 0$  means that no topological constraint is performed. When  $\alpha$  increases, the key performance indicator F1 score, which considers both precision and recall, first increases slowly and then decreases sharply. Since the F1 score reaches the maximum when  $\alpha = 0.6$ , we fix it as 0.6 in the following experiments.

The similarity threshold of images  $\beta$  determines whether a test image contains any logo or not. The experimental results for different values of  $\beta$  are shown in Fig. 4(b) with  $\alpha = 0.6$  and the number of selected features equal to 100. As we can see, with the increasing value of  $\beta$ , the F1 score first increases and then decreases. When  $\beta$  reaches the value of 4, F1 score and the precision are both comparatively high, so we fix it as 4 in the following experiments.

The third important parameter is the number of selected features. Figure 4(c) and (d) show how it affects the performance of F1 score, as well as the speed of recognition, respectively. With the increasing number of selected features, the F1 score first increases and then decreases sharply, and the time cost for recognition keeps increasing. We found that the selected features with number 100 give the best tradeoff between F1 score and the time cost. So in the following experiments, we fix the number of selected features as 100.

## 4.2 FlickrLogos-32 Dataset

Romberg et al. [4] proposed a Bundle min-Hashing (BmH) approach for logo recognition which achieves the best result of FlickrLogos-32 so far. They used



**Fig. 5.** An illustration of the process how the proposed approach filters out the mismatched keypoints between (a) and (b) and keeps the truly matched keypoints between (c) and (d), resulting in getting more accurate recognition results.

the DoG detector and the SIFT descriptor, and employed Approximate K-Means (AKM) to quantize the descriptor vectors to visual words. For fair comparisons, we also use the 1 M dimension BoW features they provided in our experiment.

Besides Bundle min-Hashing (BmH), we report the comparison results of the proposed approach with several common methods including Scalable Logo Recognition (SLR) [3], Tree-based Shape Descriptor (TSD) [1] and Correlation-Based Burstiness (CBB) [2] in terms of F1 score. In order to verify the necessity and effectiveness of each step in our approach, we design three baselines: Baseline1 does not use feature selection or any constraint, Baseline2 only uses feature selection and Baseline3 uses feature selection and the constraint proposed by Tell et al. [12]. The difference between ITC and Baseline3 is that every three keypoints of the cyclic LCS obtained by ITC satisfy Lemma 1. As shown in Table 1, the result of ITC outperforms the state-of-the-art result of BmH. Moreover, each step of our approach does make certain contribution to the final result.

In Fig. 5, we show how the proposed approach filters out the mismatched keypoints between Fig. 5(a) and (b), and keeps the truly matched keypoints between Fig. 5(c) and (d). Since there are many similar keypoints around the letters in Fig. 5(a) and (b), the matched keypoints between them are more than the matched keypoints between Fig. 5(c) and (d). After the process of feature selection, about half of the matched keypoints between Fig. 5(a) and (b) are filtered out since they are irrelevant to the target logo “Ritter SPORT” of Fig. 5(b). Moreover, by further applying the topological verification with the improved topological constraint, the matched keypoints between Fig. 5(a) and (b) are all removed because they have different spatial distribution. Finally, the number of matched keypoints between Fig. 5(c) and (d) becomes larger than the number of matched keypoints between Fig. 5(a) and (b), thus the images in Fig. 5(a) and (c) which contain the logo “ALDI” will be recognized as “ALDI” rather than “Ritter SPORT”.

**Table 1.** Performance of ITC (the proposed approach) against SLR, TSD, CBB, BmH and three baselines on FlickrLogos-32.

Method	SLR	TSD	CBB	BmH	Baseline1	Baseline2	Baseline3	ITC
<b>Precision</b>	0.980	0.980	0.980	0.999	0.957	0.992	0.994	0.994
<b>Recall</b>	0.610	0.680	0.730	0.832	0.023	0.782	0.821	0.868
<b>F1</b>	0.752	0.803	0.837	0.908	0.045	0.875	0.899	<b>0.927</b>

**Table 2.** Performance of ITC (the proposed approach) against msDT, TSD, Baseline1, Baseline2 and Baseline3 on FlickrLogos-27.

Method	msDT	TSD	Baseline1	Baseline2	Baseline3	ITC
<b>Accuracy</b>	0.520	0.580	0.500	0.542	0.619	<b>0.637</b>

### 4.3 FlickrLogos-27 Dataset

Following the experimental settings in Kalantidis et al. [1], we apply the SURF descriptors as local features. Then a vocabulary of 5K visual words is built to quantize all these descriptors. We compare the proposed approach with msDT [1], TSD [5] and the three baselines. TSD achieves the state-of-the-art result on FlickrLogos-27 so far when the distractor set is used. The comparison results are shown in Table 2, in terms of accuracy as in Kalantidis et al. [1].

We can see that, our best result is about 6% higher than the state-of-the-art result of TSD. As same as the result on FlickrLogos-32, each step of our approach does make certain contribution to the final result, suggesting the proposed approach is steady when logo classes increase to several thousands.

## 5 Conclusion

In this paper, we have proposed an Improved Topological Constraint (ITC) and a new logo recognition approach which combines both feature selection and ITC. Firstly, feature selection is used to filter out most keypoints which are irrelevant to the target logo. Then ITC is used to reduce the number of mismatching by considering relative position between a keypoint and its neighboring points. The experimental results on challenging logo datasets have shown the effectiveness of our approach.

**Acknowledgements.** This work was supported by National Natural Science Foundation of China under Grants 61371128 and 61532005, and National Hi-Tech Research and Development Program of China (863 Program) under Grants 2014AA015102 and 2012AA012503.

## References

1. Kalantidis, Y., Pueyo, L. G., Trevisiol, M., Van Zwol, R., Avrithis, Y.: Scalable triangulation-based logo recognition. In: ACM International Conference on Multimedia Retrieval (ICMR), p. 20. ACM (2011)
2. Revaud, J., Douze, M., Schmid, C.: Correlation-based burstiness for logo retrieval. In: ACM International Conference on Multimedia (ACM-MM), pp. 965–968. ACM (2012)
3. Romberg, S., Pueyo, L.G., Lienhart, R., Van Zwol, R.: Scalable logo recognition in real-world images. In: ACM International Conference on Multimedia Retrieval (ICMR), p. 25. ACM (2011)
4. Romberg, S., Lienhart, R.: Bundle min-hashing for logo recognition. In: ACM International Conference on Multimedia Retrieval (ICMR), pp. 113–120. ACM (2013)
5. Wan, C., Zhao, Z., Guo, X., Cai, A.: Tree-based shape descriptor for scalable logo detection. In: Visual Communications and Image Processing (VCIP), pp. 1–6. IEEE (2013)
6. Romberg, S.: From local features to local regions. In: ACM International Conference on Multimedia (ACM-MM), pp. 841–844. ACM (2011)
7. Wu, X., Kashino, K.: Image retrieval based on spatial context with relaxed gabriel graph pyramid. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6879–6883. IEEE (2014)
8. Zhou, W., Lu, Y., et al.: Spatial coding for large scale partial-duplicate web image search. In: ACM International Conference on Multimedia (ACM-MM), pp. 511–520. ACM (2010)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis. (IJCV)* **60**(2), 91–110 (2004)
10. Gregor, J., Thomason, M., et al.: Dynamic programming alignment of sequences representing cyclic patterns. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **15**(2), 129–135 (1993)
11. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*, vol. 1. Cambridge University Press, Cambridge (2008)
12. Tell, D., Carlsson, S.: Combining appearance and topology for wide baseline matching. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002, Part I. LNCS*, vol. 2350, pp. 68–81. Springer, Heidelberg (2002)