

INDIVIDUALIZED MATCHING BASED ON LOGO DENSITY FOR SCALABLE LOGO RECOGNITION

Yuan Zhang, Shuwu Zhang, Wei Liang and Qinzhen Guo

Institute of Automation, Chinese Academy of Sciences, Beijing, China
{yuan.zhang, shuwu.zhang, wei.liang, qinzhen.guo}@ia.ac.cn

ABSTRACT

Although many systems based on global or local descriptors have shown promising results for logo recognition, they have handled all logos with the same structure and not considered their diversities. Therefore, with the logo scale increasing, the general way cannot recognize each logo perfectly. To overcome this limitation, we propose a novel strategy to match query and each logo individually using these features. First, a new conception named logo density is introduced as important semantic information for logos. Second, matching density is given according to the logo density and by utilizing it in logistic function an individualized matching strategy is developed to obtain accurate similarity for query and a logo. Finally, we present a fast recognition algorithm based upon bag-of-words model to realize scalable logo recognition. Our method is evaluated on two challenging datasets (our 10,000-class logo dataset and FlickrLogos-27). Experiments demonstrate its superior performance comparing to previous methods.

Index Terms— individualized matching, logo density, logistic function, scalable logo recognition

1. INTRODUCTION

One challenging work of computer vision is specific object recognition. Meanwhile, with the explosive growth of object categories, the demand for real-time object recognition in the large scale database is becoming increasingly high, especially certain region recognition in mobile application and recommender system. In this paper we focus on logo recognition. Logos, a symbol of identification for companies, products and organizations, can be regarded as objects with a planar surface and are significantly valuable in many practical scenarios of modern marketing, advertising and trademark registration [1]. Typically, the database of logos can scale up to thousands of classes. Such a scale is not very common in generic object detection or recognition [2]. Our goal is that given a query region or an unknown image in natural scenes and a large logo database, recognize the logo contained in the query in real-time, if any (shown in Fig. 1).

Previous research has mainly paid attention to logo recognition and classification in text documents [3-8]. However, in these cases, the logos are achromatic, frontally



Fig. 1 Examples of logos displayed in real world images: Pepsi, Texaco, Starbucks and Unilever. The left two logos are simple while the right two are complex.

viewed and positioned on clean backgrounds, because of which they cannot be applied to real world images. In content-based logo retrieval, shape is particularly relevant when it comes to assess the logo identity, thus shape descriptors are more often used for logo recognition, such as shape context (SC) in [9]. Then Mori *et al.* constructed a richer descriptor than SC based on oriented edges, called generalized shape contexts (GSCs) [10]. Several authors have proposed the combination of color and shape for logo detection in real world images [11-13].

Interest points and local descriptors appear much more appropriate to support detection and recognition of graphic logos in real world images, like SIFT [14] and SURF [15]. Recently, many researchers have extended them with spatial information of logos. Sahbi *et al.* [16] introduced a novel logo detection and localization approach based on a new class of similarities referred to context dependent. Romberg *et al.* [17] developed a highly effective and scalable framework to encode and index the relative spatial layout of local features. Kalantidis *et al.* [2] also proposed a scalable logo recognition approach that extended the common bag-of-words model and incorporated local geometry in the indexing process.

Despite inspiring results of these researches, it should be noted that they have dealt with all kinds of logos within the same framework. When the logo database scales up to thousands of classes, these features are semantically too low-level to substantially describe such a large variety. For methods with shape-based features, when the boundary of a shape is not a whole closed curve but a curve consisting of several closed ones such as Starbucks and Unilever in Fig.1, we couldn't obtain accurate shape features using the continuous feature descriptors. Therefore, when logo shape is complicated and the whole boundary information is

difficult to get, the boundary-based descriptors are not suitable. For systems with interest point descriptors like SIFT, they cannot detect sufficient keypoints due to large smooth regions, degradations and simplicity in many logos (e.g., Pepsi and Texaco in Fig.1). As a result, these descriptors may be somewhat ineffective and lead to low recall and precision for simple logos [2, 16 and 17].

Consequently, considering such a large variety of logos, we focus on developing a novel logo recognition algorithm which deals with each logo individually and effectively in line with semantic information. We first propose a new idea in view of the number of interest points to measure logo density which is not mentioned before, then we present an individualized matching strategy (IMS) to combine SIFTs and GSCs based upon the density so that these features can model the logo best. Finally in order to tackle logo recognition in large scale database, we use the bag-of-words model [18] as in [2, 17] and give the whole algorithm. Our method can offset limitations of shape descriptors and local descriptors to perform scalable logo recognition and guarantee both high precision and efficiency for various logos. Moreover, the proposed method is training-free.

The rest of our paper is organized as follows. Section 2 shows the definition of logo density and individualized matching strategy. The scalable logo recognition algorithm is given in Section 3. The experimental results are presented in section 4 and final conclusions are given in section 5.

2. INDIVIDUALIZED MATCHING

2.1. Logo Analysis

Logos are graphic productions. Some are made up of simple patterns, such as “Apple”, “Pepsi”, and “Mc Donald’s”. Some are composed of plenty of different regions, like “Unicef”, “Nestle” and “Unilever”. The complicated logos contain much richer information than the simple one. And the more complicated a logo is, the more accurate local descriptors are. On the contrary, the simpler a logo is, the better global descriptors perform. Thus these two features can be mutually complementary and reinforcing. Combining them in a suitable way, we will achieve more robustness with less effort to recognize various logos successfully.

2.2. Logo Density

While combining these two features, we plan to assign different proportions for them according to the complexity of each logo. It is necessary to establish a criterion to measure how complicated the logo is. In this paper, we put forward a new conception named logo density. Like the literal meaning of “density”, it is the property of a logo which can show its complexity. Our research has found that the number of SIFT keypoints can reflect it to some extent. Complicated logos have more keypoints than simple ones in the same image size. Hence we give its definition as follows:

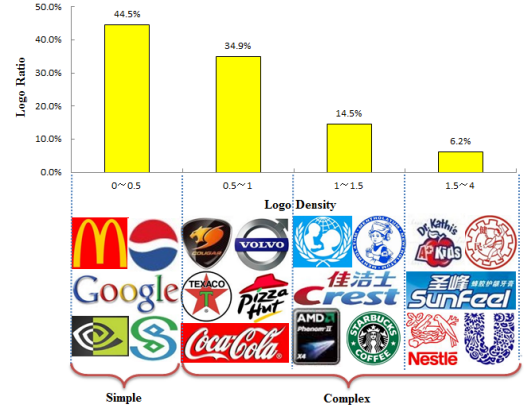


Fig. 2 Logo distribution based on logo density and some logos in corresponding regions

Table 1. The logo density of some logos

Logo name	ρ_L	Logo name	ρ_L
McDonalds	0.039	Cougar	0.540
Pepsi	0.174	Volvo	0.562
Google	0.294	Texaco	0.645
Nvidia	0.358	PizzaHut	0.735
CMBC	0.465	Cocacola	0.824
Logo name	ρ_L	Logo name	ρ_L
Unicef	1.102	AKIDs	1.611
Mentholatum	1.274	Wuhan Jianming	1.753
Crest	1.328	SunFeel	1.787
AMD	1.398	Nestle	1.889
Starbucks	1.461	Unilever	2.973

Definition: Logo density (ρ_L) is the average number of keypoints in a unit area (UA), as shown in (1) and (2):

$$\rho_L = \frac{KN}{UAN} \quad (1)$$

$$UAN = \frac{Area}{UA} \quad (2)$$

where KN is its number of keypoints and UAN refers to the number of UA in the logo image. KN and UAN can be seen as “mass” and “volume” of a logo respectively. Here we set UA to 100. Area is the product of height and width of the image.

As shown in Fig. 2, we draw the distribution of ρ_L for 10,000 kinds of logos in our dataset. Meanwhile, we give some examples in corresponding regions and their densities in Table 1. It is obvious that as the ρ_L increases, logos become increasingly complex. Here a logo with its ρ_L bigger than 0.5 is considered to be complex (“heavy”), otherwise it is simple (“light”). Therefore, logo density can provide some semantic information to realize perfect combination of the two features. Fig. 2 illustrates that the number of simple logos is almost equal to that of complex ones, which means combining the two features together can lead to a comprehensive result for logo recognition.

2.3. Individualized Matching Strategy

In this paper, we employ SIFT features as local descriptors for its capability of capturing sufficient discriminative local elements and GSCs as the global descriptors for its strength

of outstanding shape representation. Two similarities are obtained using SIFT descriptors and GSCs between a query and a logo, and then combined as follows:

$$Sim(Q_{L_i}) = wSim_{SIFT}(Q_{L_i}) + (1 - w)Sim_{GSCs}(Q_{L_i}) \quad (3)$$

where w is the weight and $0 \leq w \leq 1$. $Sim(Q_{L_i})$ is the final similarity between the query Q and logo L_i which is the i_{th} logo in the dataset of N logos, $i \in \{1, \dots, N\}$. $Sim_{SIFT}(Q_{L_i})$ and $Sim_{GSCs}(Q_{L_i})$ are similarities using SIFT and GSCs correspondingly.

As logos have great diversities, we need to adaptively select the weight w for each logo. Hence we will highlight the novel idea of adaptive weight based on the logo density. As mentioned before, the more complicated a logo is, the more accurate local descriptors are. On the contrary, the simpler a logo is, the better global descriptors perform. So the logo density can be taken as a variable to compute weights for the two situations, which can be considered as the binary classification to some extent. Thus we employ logistic regression and adjust it to make it suitable for our approach as follows:

$$w = f(\rho_M - T_\rho) = \frac{1}{1 + e^{-(\rho_M - T_\rho)}} \quad (4)$$

$$\rho_M = \alpha \cdot \rho_Q + \beta \cdot \rho_{L_i} \quad (5)$$

where ρ_M is the matching density. For Q , ρ_Q decides which feature is more reliable between SIFT and GSCs. And so does ρ_{L_i} for L_i . Thus we should take both ρ_Q and ρ_{L_i} into consideration and combine them as ρ_M in (5). In this paper, we reckon that they are both important and set $\alpha = \beta = 0.5$. When ρ_M is low, we name the match “simple match”. When ρ_M is high, we name it “complicated match”. In other cases, we deem it as “median match”. In (4) $f(\bullet)$ is the logistic function with its value between 0 and 1. The output can be interpreted as the weight w in (3). $\rho_M - T_\rho = 0$ is the decision boundary, and we have $w = f(0) = 0.5$ which means for this match between Q and L_i , both SIFT and GSCs are equally important. With $\rho_M - T_\rho$ increasing, w becomes bigger and bigger which implies SIFT is more and more effective than GSCs. Contrarily, with $\rho_M - T_\rho$ decreasing, w becomes smaller and smaller which signifies GSCs is more and more effective than SIFT. In our approach, T_ρ , as the density threshold, is estimated from validation dataset while getting the best performance.

Therefore through IMS, we adopt feature preference of query and each logo into their match based on logo density so that we can treat each match individually and gain more accurate similarity for them.

3. SCALABLE LOGO RECOGNITION ALGORITHM

A turning point in scalable image object retrieval, which also provides the basis for our approach is the introduction of the bag-of-words (BOW) [18]. At the core of the visual BOW is the quantization of local features for efficient

indexing instead of matching the raw descriptors directly. So in this work we use SIFT and GSCs descriptors to build two inverted indexes separately for our 10,000 classes of logo images, each with 10K visual words. When forming vocabularies, we employ HKM [19] clustering which deals with a large scale database effectively. Based on the two indexes we give our logo matching steps in *Algorithm 1*.

Algorithm 1 Scalable Logo Recognition for a Query

Step 1: Match the query with BOW-SIFT inverted index and BOW-GSCs inverted index respectively. Then get two similarity ranked lists in descending order for logos.

Step 2: For the top R (e.g. 20) logos in BOW-SIFT ranked list, obtain corresponding $Sim_{GSCs}(Q_{L_i})$ in BOW-GSCs ranked list and compute $Sim(Q_{L_i})$ for them using (3). So do the top R logos in BOW-GSCs ranked list. Then achieve the combination list of $2 \times R$ logos at most.

Step 3: Reorder the combination list in descending order by $Sim(Q_{L_i})$. If the first logo's similarity of the list is bigger than threshold T_{sim} , return it as the recognition result for the query. Otherwise, the query does not contain a logo.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Our Dataset and Setup

For a realistic evaluation of our proposed approach, a large collection of images in a real world environment is required. Therefore a total of 24 classes were chosen for the dataset, each one corresponding to a brand. We selected 1,700 images for all classes from Flickr and Google, so that every image contained one instance of the brand's logo. We manually annotated all images with bounding boxes for the logo instances as the query region. Logos appear with rotation, spatial distortion, difficult illumination, scaling, etc. The annotated collection of logos was then split into two disjunct subsets P_1 and P_2 , each containing images of all 24 classes. In total, both P_1 and P_2 have 850 annotated images. In our experiments we choose P_1 as validation set to tune the parameters and P_2 as test set to report performance. To complete both sets with negative images that do not have any logos, we gathered 200 images taken in natural environments, without any logos for P_1 and P_2 respectively.

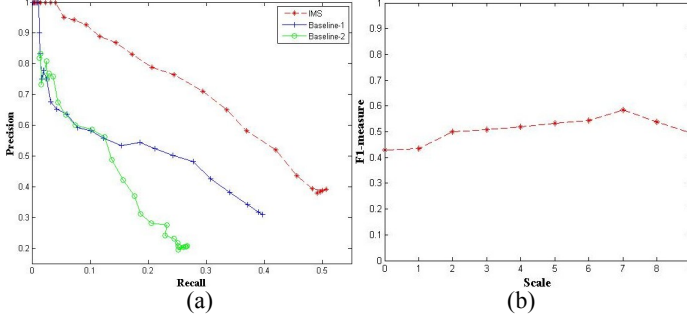
Then, we crawled logos of 10,000 different classes from Flickr and Google as the logo database, with just one image per class. These logos are of different degrees of complexity and adjusted to a medium resolution (158×158). We ensure that all logos appear clear and upright in the images.

In order to be robust against appearance variations, two visual codebooks of 10K visual words are used for SIFT and GSCs respectively. The purpose is to recognize which logo the query is from the 10,000 classes. And we compare our approach with two state-of-the-art methods, referring BOW-SIFT as baseline-1 and BOW-GSCs as baseline-2.

4.2. Experiments on Our Dataset

Table 2. The precision (P), recall (R) and f1-measure (F1) for five samples and all logos in the test dataset using IMS, B-1 and B-2

Category	Bank of China(0.113)			Pepsi(0.174)			Mengniu(0.368)			YILI(0.643)			Starbucks(1.461)			All		
(ρ_L)	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
IMS	0.94	0.77	0.85	0.43	0.19	0.27	0.71	0.40	0.51	0.75	0.60	0.67	0.89	0.72	0.79	0.52	0.45	0.48
B-1	0.63	0.59	0.61	0.03	0.02	0.02	0.40	0.24	0.30	0.63	0.50	0.56	0.73	0.66	0.69	0.35	0.33	0.34
B-2	0.61	0.60	0.61	0.24	0.23	0.24	0.36	0.36	0.36	0.60	0.60	0.60	0.13	0.13	0.13	0.21	0.28	0.24

**Fig. 3.** (a) Precision & Recall curves for validation set, (b) Scale & F1-measure curve

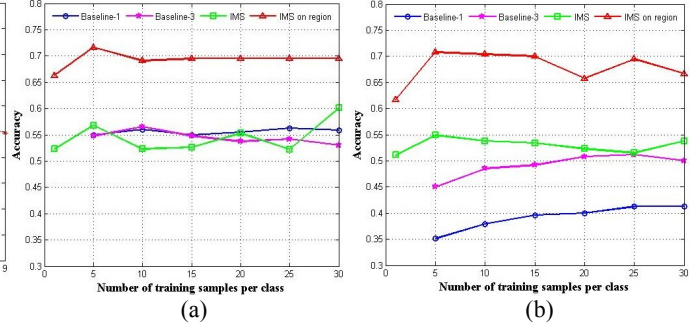
In our model, one crucial parameter is T_ρ in (4). We vary T_ρ between 0 and 1 and get the performance in terms of precision on positive examples in P_1 with $T_{sim} = 0$. When T_ρ is 0.15, the precision is the highest. So we set $T_\rho = 0.15$. Then we give the precision & recall curves for our method (IMS), baseline-1 (B-1) and baseline-2 (B-2) in Fig. 3(a), which indicates that our proposed approach outperforms the two baselines in all precision and recall settlements. We determine the optimal $T_{sim} = 0.18$ when obtaining the highest F1-measure.

After gaining all parameters through P_1 for our method, Table 2 presents the precision and recall for five samples with different density in P_2 which further underlines the better effectiveness of our logo recognition system. And we have an average precision of 0.52 and recall of 0.45 for all 24 logo classes. The obtained results clearly show that our approach produces very good performance for all logos, both simple and complicated, while ensuring an average speed of 0.452s per query. We believe that the results will be better with more images per class in our logo database.

In order to show the robustness of our density based method to image size, we make another experiment on different scales of images and show precisions in Fig. 3(b). Scale- i means to shrink original image by 0.9^i . It demonstrates that image scale does not have a major impact on performance.

4.3. Experiments on FlickrLogos-27

We also report results on the FlickrLogos-27 image collection, to demonstrate the generality of our method. Their paper [2] reported results obtained using a common BOW model (baseline-1) vs their msDT approach (baseline-3). Performances are reported in terms of accuracy by varying the number of training images per class. Since our method does not provide a learning phase, we use the “training images” as reference logos. Therefore, if we have k training images, we set k images for one class in the logo dataset and finally we assign to each query image the label

**Fig. 4.** Performance of our approach versus baseline-1 and baseline-3 [2] on the FlickrLogos-27 dataset (query regions and the whole images) with only 27 classes present in the index (a) and more than 4K classes in total in the index (b).

corresponding to the reference image that maximizes criterion (3). We annotate query regions in its test set and report the results for both query regions and the whole images compared to baseline-1 and baseline-3 on the indexes of 27 and 4K classes with 5K and 10K visual words in Fig. 4(a)&(b). As demonstrated by this figure, our method guarantees very good performance also using a single reference logo and substantially outperforms both methods with more reference images for query regions and the whole images of all logo classes in different scales. Meanwhile, we achieve a fast speed of 0.74s for query region and 3.4s for the whole images on the index of 4K classes.

5. CONCLUSION

In this paper, we have proposed a novel and effective criterion to measure logo density and presented our individualized matching strategy based on this density to treat each match individually. Then we give our scalable logo recognition algorithm combined with this strategy to realize fast matching. Experiments on our large dataset and FlickrLogos-27 dataset show that our system can ensure more accurate recognition for both simple and complicated logos than the state-of-art methods. Moreover, we achieve fast logo recognition in large logo dataset. In future work, we are interested in logo detection algorithm to obtain possible logo regions like face detection. In addition, we would study more proper features to recognize various logos.

6. ACKNOWLEDGEMENTS

The work is supported by the National Key Technology R&D Program of China under Grant No.2012BAH04F02, 2012BAH88F02, 2013BAH61F01 and 2013BAH63F01, and the International S&T Cooperation Program of China under Grant No.2013DFG12980.

7. REFERENCES

- [1] S.Y. Arafat, S.A. Husain, I.A. Niaz and M. Saleem, "Logo Detection and Recognition in Video Stream," *ICDIM*, pp. 163-168, 2010.
- [2] Y. Kalantidis, L. G. Pueyo, M. Trevisiol, R. van Zwol, and Y. Avrithis, "Scalable triangulation-based logo recognition," in *Proc. ACM Int. Conf. Multimedia Retr.*, Trento, Italy, 2011, pp. 1-7.
- [3] D.Doermann, E.Rivlin and I.Weiss, "Logo recognition using geometric invariants," *Procee International Conference on Document Analysis and Recognition*, pages 894-897, 1993.
- [4] J.Neumann, H.Samet, and A.Soffer. "Integration of local and global shape analysis for logo classification," *Pattern Recognition Letters*, 23(12), pages1449-1457, 2002.
- [5] G.Zhu and D.Doermann, "Automatic Document Logo Detection," *International Conference on Document Analysis and Recognition*, pages 864-868, 2007.
- [6] H.Wang and Y.Chen, "Logo Detection in Document Images Based on Boundary Extension of feature rectangles," *International Conference on Document Analysis and Recognition*, pages 1335-1339, 2009.
- [7] Zhe Li, Matthias Schulte-Austum and Martin Neschen, "Fast Logo Detection and Recognition in Document Images," in *ICPR*, pages 2716-2719, 2010.
- [8] Hongye Wang, "Document Logo Detection and Recognition Using Bayesian Model," in *ICPR*, pages 1961-1964, 2010.
- [9] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509-522, Apr. 2002.
- [10] G. Mori, S.Belongie and J.Malik, "Efficient Shape Matching Using Shape Contexts," *IEEE Trans. PAMI*, 27(11), 2005.
- [11] J. Luo and D. Crandall, "Color object detection using spatial-color joint probability functions," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1443-1453, Jun. 2006.
- [12] R. Phan, J. Chia, and D. Androutsos, "Unconstrained logo and trademark retrieval in general color image databases using color edge gradient co-occurrence histograms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Las Vegas, NV, Mar.-Apr. 2008, pp. 1221-1224.
- [13] R. Phan and D. Androutsos, "Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms," *Comput. Vis. Image Understand.*, vol. 114, no. 1, pp. 66-84, 2010.
- [14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91-110, 2004.
- [15] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346-359, 2008.
- [16] H. Sahbi, L. Ballan, G. Serra, and A. Del Bimbo, "Context-Dependent Logo Matching and Recognition," *IEEE Trans. Image Process.*, vol. 22, pp. 1018-1031, March. 2013.
- [17] S.Romberg, L.G. Pueyo, R.Lienhart, and R.V. Zwol, "Scalable logo recognition in real-world images," *Proc. of ICMR*, 2011.
- [18] J. Sivic and A. Zisserman. "Video google: A text retrieval approach to object matching in videos," In *Proceedings of the 9th International Conference on Computer Vision*, Nice, France. pages 1470-1478, 2003.
- [19] D. Nister and H. Stewenius. "Scalable recognition with a vocabulary tree," In *Proc. CVPR*, vol. 2, pp. 2161-2168, 2006.