Logo Localization and Recognition Based on Spatial Pyramid Matching with Color Proportion

Wenya Feng[1], Zhiqian Chen, Yonggan Hou, Long Liu, Yilin Guo

Abstract. In this paper, we propose a novel logo localization and recognition model in which we use the spatial pyramid matching with color proportion to train a score function and the Efficient Subwindow Search (ESS) to localize region and recognize specific logo. Our model has two advantages compared with other related algorithms: (1) this model has a $O(n^2)$ time complexity in localizing logo regions which is much faster than the traditional sliding window approaches that has a time complexity of $O(n^4)$ which make it cannot apply to real world applications because of the high computational cost; (2) by means of our well-designed spatial pyramid matching embedded by color proportion, we can achieve a higher recognition rate even with the linear kernel SVM. We test our model in a logo dataset called FlickrLogos-32 which contains 32 different logo classes by downloading them from Flickr. Experiments show that our model both have fast enough speed and high correct rate.

Keywords: Logo Localization, Logo Recognition, Efficient Subwindow Search, Color Proportion

1 Introduction

The most common and challenge work in computer vision is certainly object recognition. Although recent years have seen great progress in object recognition field, nevertheless, due to many difficulties such as object size, light, location, different views etc, object recognition is still a hard problem. Logos are widely used visually-salient symbols serving as remarkable identifications of related organization. So logo recognition can result in very wide application.

Because logos are designed for representing companies in some way, different company usually has different logo. In addition, in order to make their company easy to remember and identify, companies always create especially symbols as

---

[1] Wenya. Feng (✉), Zhiqian Chen, Yonggan Hou, Long Liu, Yilin Guo
Peking University, Beijing, P.R. China
e-mail: fengwenya@pku.edu.cn

logo. So logos have different geometry structure and different color composition. This make the multi-logo localization and recognition extremely challenge. In this paper, we propose an efficient algorithm which use Spatial Pyramid Matching with Color Proportion (SPMCP) and Efficient Subwindow Search(ESS) to complete the localization and recognition problems simultaneously and because of the low time complexity this algorithm can be used in real time logo detection application.

2 Related Work

Many different form definition of object location exist in literature, e.g. by its contour, center point, a segment, or a bounding box. In this paper, we use the bounding box method to represent a detected logo region. Techniques in logo detection mainly include localization and recognition. Localization is the most time-consuming part in this algorithm; recognition mainly depends on the feature and the classifier.

2.1 Logo Localization

Logo localization is an important task for automatic understanding images and to separate a logo from background in real-world logo detection. Sliding window approaches have established themselves as state-of-the-art for many years [1-5]. This kind of algorithms treats localization as localized classification. Applying a classifier function subsequently to subimages within an image and taking the region with the maximum score as the logo region. But even a 200x200 image contains more than one billion rectangular subimages, and the number of subimages grows quadratically with the number of pixels which makes it computationally too expensive to evaluate the score function exhaustively for all of subimages. Many researcher have developed algorithm to reduce the number of evaluation of score function by only search certain fixed sizes as candidates or a coarse grid of possible location[8][10]. Above reduced search techniques sacrifice localization to fast algorithms' speed, which may lead to false estimations or even complete misses of logos. In this paper, we localize logo by the efficient subwindow search (ESS) [6] which is a simple yet powerful branch and bound scheme that allows efficient maximization a large class of score function over all possible subimages. At the same time ESS is very fast, because it relies on a branch-and-bound search instead of an exhaustive search. This speed advantage allows the use of more complex and better classifiers.

2.2 Logo Recognition

Compared with localization problem, logo recognition is a relative easy issue. [7-9] used neural networks to recognize localized logos; [10] created feature vector by means of SIFT to recognize vehicle logo; [11] used Spatial Pyramid Matching to detect logo in natural scenes. Spatial Pyramid Matching [12, 13] can be considered as bag of visual word[14] with location information. In this paper, we design a novel Spatial Pyramid Matching with Color Proportion (SPMCP) which fully makes use of the color characteristic that are usually a remarkable feature of logo design purpose and we use the linear kernel to recognize logos. One thing need to be emphasized is that our logo recognition phase is combined with our logo localization phase which means when the logo is localized the logo will be recognized at same time.

3 Color Proportion Algorithm

	L*	a*	b*		L*	a*	b*
1	38	57	36	20	68	31	-13
2	44	63	38	21	52	58	17
3	54	81	70	22	65	40	20
4	64	51	44	23	74	27	12
5	42	-40	24	24	88	15	10
6	64	-44	22	25	28	4	-40
7	75	-28	15	26	30	68	-112
8	88	-79	81	27	49	-4	-42
9	17	28	11	28	62	-6	-35
10	31	21	26	29	63	48	62
11	47	18	45	30	67	40	68
12	66	20	61	31	75	26	76
13	93	-8	74	32	85	9	67
14	95	-7	51	33	0	0	0
15	97	-4	29	34	20	0	0
16	98	-16	93	35	40	0	0
17	30	46	-24	36	60	0	0
18	40	59	-30	37	80	0	0
19	52	48	-19	38	100	0	0

Fig. 1 Experimental results of 38 colors in the L*a*b* color space

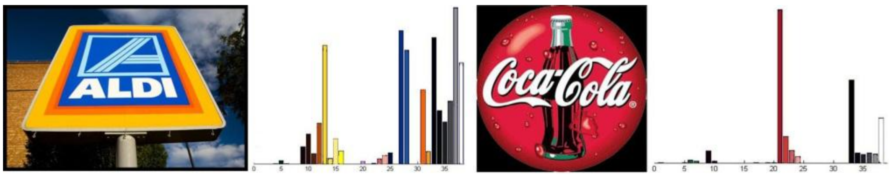
Recent years, bag of word methods become very popular in image retrieval, object recognition area. These kinds of algorithms usually involve some feature extraction techniques e.g. SIFT, SURF, HOG which pay no attention or give less weight on object's color characteristic [6, 13]. However, color proportion plays a fundamental role in image retrieval, object recognition related applications, especially in logo detection algorithms. In this section, we will introduce our well-designed color proportion algorithm devised in the L*a*b* color space with efficacious color distance [15]. The reason of using the L*a*b* color space is that the color space has special property that luminance and chromatic information are represented separately by which we can divide red, crimson, light red into same class. This function is very useful in logo detection, for example an image with a red logo may be converting to a crimson logo. According to [14] colors can be divided into 3 main categories: primary colors, secondary colors and complementary colors. All other colors are made out of these colors (e.g. the mixture of red and blue makes purple; the mixture of yellow and blue makes green).

With our experimental results of 38 colors (Fig. 1) and the following color distance, we can generate a 38 bins histogram to represent image's color proportion:

$$\begin{cases} d_L = I_i(L) - I_j(L) \\ d_a = I_i(a) - I_j(a) \\ d_b = I_i(b) - I_j(b) \\ D(I_i, I_j) = \sqrt{w_L \cdot d_L^2 + w_a \cdot d_a^2 + w_b \cdot d_b^2} \\ s.t. \quad w_L + w_a + w_b = 1, 0 < w_L, w_a, w_b < 1 \end{cases} \qquad (1)$$

where $D(I_i, I_j)$ represents the distance between pixel i and j in an image; $i, j \in \Omega \subset \mathbb{R}^2$ denotes the spatial domain, $I_i \in \Gamma \subset \mathbb{R}^3$ denotes a discrete color image; image I in the $L^*a^*b^*$ color space define as $I : I_i = (I_i(L), I_i(a), I_i(b))^T$; $w_L : w_a : w_b = 1:7:7$ Similar to [16], the weight of L channel is reduced compared with *a, b* channels, therefore the distance between similar chromatic colors with different luminance is decreased. Hence, similar colors become easier to be assigned to same categorization even if they have different luminance.

With the above distance and the experiment results in Figure 1 we can generate an image's color proportion representation by a 38 bins histogram (Figure 2).



(a) (b) (c) (d)

Fig. 2 (a) (c): logos come from FlickrLogos-32 dataset; (b)(d) their corresponding 38 bins histograms obtained by our method

4. Localization by Spatial Pyramid Matching with Color Proportion (SPMCP)

4.1 Score Function

In order to find an exact logo region in an image, we design a score function by which the region with the highest score indicates a candidate logo region. The score function should subject to the following constrain: (1) the exact logo region should has the highest score; (2) a bigger region contains the exact logo region have smaller score; (3) other regions which does not include the logo region have smaller score (usually minus score) than the regions contain logo (Figure 3)

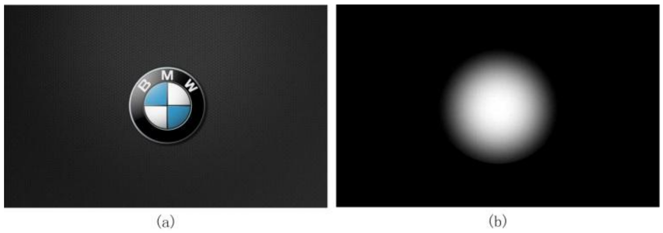


Fig. 3 (a) An example image with BMW logo; (b) its corresponding feature map in which white represents positive score and black represents negative score

We call the images like Figure 3(b) as feature map which subject to above three constrains. The feature maps are calculated by our score function. As we all know, raw bag of words models have no notion of geometry. They are therefore not the best choice for the detection of logo classes which have characteristic geometric arrangements. Spatial pyramid features [13] have been developed to overcome this limitation. They divide every image into a grid of patches and represent each grid cell by a separate histogram. Here, we use SIFT to exact feature of every patches. Under this situation, every patch has a 128 dimensional vector. Because the SIFT feature is derive from gray image which lose the color attribute and even though we exact feature by other feature exaction algorithms their color attribute are not as good as our well-designed color proportion algorithm introduced in Section 3. Hence, when adding our 38 dimensional color histogram to the 128 dimensional feature vectors will let the feature vector have the ability of representing both texture structure and color characteristic. We consider the linear kernel SVM classifier with on top of such a hierarchical spatial pyramid histogram representation. The decision function *f* for a region *x* in an image is calculated as:

$$f(x) = \beta + \sum_{i=1}^{L} \sum_{\substack{i=1...l \\ j=1...l}} \sum_{k=1}^{N} w_k^{l,(i,j)} h_{i,j}^k \qquad (2)$$

Where $h_{i,j}^k$ is the histogram of all feature that fall into the spatial pyramid grid cell with index (*i,j*). β and $w_k^{l,(i,j)}$ are the coefficients learned by the linear SVM when trained with the 166 bins histogram vector and we can set the bias term β as 0. By means of the linearity of the scalar products, we can transform Eq(2) into a sum of per-point contributions:

$$f(x) = \beta + \sum_{m=1}^{n} \sum_{l=1}^{L} \sum_{\substack{i=1...l \\ j=1...l}} w_{c_m}^{l,(i,j)} \qquad (3)$$

Typically, the size of region *x* usually equal to the aforementioned patch and in this case *f(x)* indicates the score of the patch *x*. Here, *f(x)>0* represents patch *x* labeled by 1, which means this patch is likely to be a part of the logo region. In contrast, $f(x) < 0$ means patch x is not likely to be a part of the logo region. We execute the $f(x)$ to all patches of an image to get a feature map like Figure 3(b). In this paper, we use rectangle to indicate detected logos and the rectangle parameterized by their top, bottom, left and right coordinates (t, b, l, r).

4.2 Branch and Bound Search

With the feature map and the score function, the next step is straightforward to construct a search schema to localize the logo target. In order to reduce the search number of patches, we build an efficient subwindow search based on branch and bound as shown in [6].

This algorithm needs two rectangles to find the location of a logo, we represent the two rectangles as R_1: $(t_1, b1, l1, r1)$, R_2: $(t_2, b2, l2, r2)$. First of all, we initialize the first rectangle R_1 as the same size of the whole image which means (t_1, l_1) indicates the left-top corner of the image and (b_1, r_1) indicates the right-bottom corner of the image (Fig.4). In contrast, we initialize the second rectangle R_2 as a minus R_1, which means (b_2, r_2) indicates the left-top corner and the (t_2, l2) indicates the right-bottom corner. We need a priority queue to store candidate rectangles and convenient to get the top as best match rectangle with the highest score. After the initialization of R_1 and R_2, we calculate the biggest interval between R_1 and R_2. Here, we define four intervals of R_1, R_2 as t_1-t_2, b_1-b_2, l_1-l_2, r_1-r_2. For example, if the term t indicate the biggest interval, we split the R_1, R_2 to two sets of rectangle (R_1', R_2'), (R_1'', R_2'') with R_1':(t_1, b_1, l_1, r_1), R_2':$((t_1+t_2)/2+1,b_2,l_2,r_2)$, R_1'':$((t_1+t_2)/2-1,b_1,l_1,r_1)$, R_2'':(t_2,b_2,l_2,r_2) and put the two sets to the priority queue. Then get the top of the priority queue to do same split

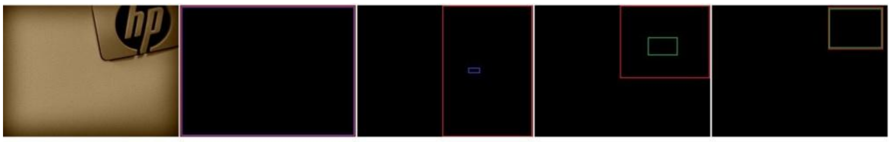


Fig. 4 (a) A sample from FlickrLogos-32 dataset. (b) Initiation of R_1, R_2, and the red rectangle represents R_1; the blue rectangle represents illegal R_2 $(t>b, l>r)$. (c) R_1 becomes closer to target and R_2 becomes closer to legal. (d) R_1 gradually reduce to logo region, the legal R_2 (green) gradually increase to logo region. (e) When R_1 and R_2 have same size, the search algorithm converges. An example image with BMW logo; (b) its corresponding feature map in which white represents positive score and black represents negative score operation until the two rectangles in same set have same size. The final two rectangles is the exact location of logo region. When apply this algorithm to multi-logo localization and recognition, one can just remove the detected logo region and restart the search algorithm to find the next logo.

5 Experimental Analysis

For the evaluation of our algorithm, we use the FlickrLogos-32[17] which consists of manually labeled logo images, complemented with nonlogo images, all posted on Flickr, as our train and test data source. We define a detected bounding box is counted as a correct match when the area of overlap with the corresponding ground truth mask box provided by FlickrLogos-32 is at least 60% of the area of their union. Figure 5 indicates precision recall plots of our results and Figure 6 indicates some detected logo samples.

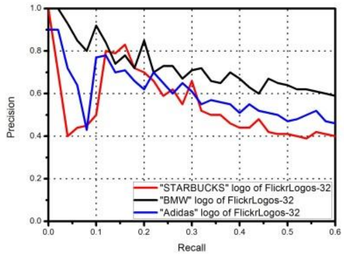


Fig. 5 Recall precision curves of three kinds of logos from FlickrLogos-32 dataset

Fig. 6 Experimental results: the green bounding box indicates detected logo region

6 Conclusion

In this paper, we proposed a logo localization and recognition algorithm based on spatial pyramid matching with color proportion. Previous researches have designed effective features [13-14] and their methods have been proved that are successful in image retrieval and object recognition area. However, those algorithms paid little attention to target's color, especially color's proportion, which can be used as helpful yet simple characteristic to achieve better performance on logo's localization and recognition. We introduce a 38 bins color histogram with well-designed color categories and distance to represent color proportion and by embedding it to the branch and bound based search schema to obtain better precise rate and faster speed.

7 References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," presented at the IEEE international conference on Computer Vision and Pattern Recognition, 2001.

[2] H. A. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes," in NIPS, 1996, pp. 875–881.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Computer Vision and Pattern Recognition, 2005., San Diego, CA, USA, 2005, pp. 886 - 893.

[4] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of Adjacent Contour Segments for Object Detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, pp. 36 - 51, Jan. 2008.

[5] O. Chum and A. Zisserman, "An Exemplar Model for Learning Object Classes," in IEEE Computer Vision and Pattern Recognition, Minneapolis, MN, 2007, pp. 1-8.

[6] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Efficient Subwindow Search: A Branch and Bound Framework for Object Localization," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, pp. 2129-2142, Dec. 2009 2009.

[7] K. Zyga, R. Price, and B. Williams, "A generalized regression neural network for logo recognition," in Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings. Fourth International Conference on, Brighton, 30 Aug 2000-01 Sep 2000, pp. 475 - 478

[8] F. Cesarini, E. Francesconi, and M. M. Gori, S. , "A neural-based architecture for spot-noisy logo recognition," in Document Analysis and Recognition., Proceedings of the Fourth International Conference on, Ulm, 18-20 Aug 1997, pp. 175 - 179

[9] E. Francesconi, P. Frasconi, M. Gori, S. Marinai, J. Q. Sheng, and A. S. G. Soda, Logo recognition by recursive neural networks: Springer Berlin Heidelberg, 1998.

[10] A. P. Psyllos, C.-N. E. Anagnostopoulos, and E. Kayafas, "Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme," Intelligent Transportation Systems, IEEE Transactions on vol. 11, pp. 322-328, June 2010.

[11] J. Kleban, X. Xie, and W.-Y. Ma, "Spatial pyramid mining for logo detection in natural scenes," in Multimedia and Expo, 2008 IEEE International Conference on, Hannover, June 23 2008-April 26 2008, pp. 1077 - 1080.

[12] K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Beijing, pp. 1458 - 1465.

[13] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in IEEE International Conference on Computer Vision and Pattern Recognition, 2006, pp. 2169 - 2178.

[14] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in IEEE Computer Vision and Pattern Recognition, 2005., 20-25 June 2005, pp. 524 - 531

[15] Wenya Feng, Yilin Guo, Okhee Kim, Yonggan Hou, Long Liu, and Huiping Sun: Color Transfer Based on Earth Mover's Distance and Color Categorization. 15th International Conference on Computer Analysis of Image and Patterns (CAIP), Aug, 2013, LNCS, Springer, York, UK.

[16] Y. Hana, X.-C. Fenga, and George Baciub, "Variational and PCA based natural image segmentation," Pattern Recognition, vol. 46, pp. 1971–1984, July 2013.

[17] Romberg, S. a. Pueyo, L. G. a. Lienhart, R. a. v. Zwol, and Roelof, "Scalable logo recognition in real-world images," in Proceedings of the 1st ACM International Conference on Multimedia Retrieval, New York, NY, USA, 2011, pp. 25:1--25:8.