14th IAPR International Conference on Machine Vision Applications (MVA)
May 18-22, 2015. Miraikan, Tokyo, Japan

METU Dataset: A Big Dataset for Benchmarking Trademark Retrieval

Osman Tursun, Sinan Kalkan
KOVAN Research Lab., Computer Engineering, Middle East Technical University
{wusiman.tuerxun, skalkan}@ceng.metu.edu.tr

Abstract

Trademark retrieval (TR) is the problem of retrieving similar trademarks (logos) for a query, and the main aim is to detect copyright infringements in trademarks. Since there are millions of companies worldwide, automatically retrieving similar trademarks has become an important problem, and currently, checking trademark infringements is mostly performed manually by humans. However, although there have been many attempts for automated TR, as also acknowledged in the community, the problem is largely unsolved. One of the main reasons for that is the unavailability of a publicly available comprehensive dataset that includes the various challenges of the TR problem. In this article, we propose and introduce a large dataset composed of more than 930,000 trademarks, and evaluate the existing approaches in the literature on this dataset. We show that the existing methods are far from being useful in such a challenging dataset, and we hope that the dataset can facilitate the development of better methods to make progress in the performance of trademark retrieval systems.

1 Introduction

A trademark is a symbol or an associated text that represents the company. For a company, a trademark can be more important than its name because it captures the nature, the philosophy and the attitude of the company.

With the open and capitalist economic models adopted by most countries in the world, more than 100 million companies are known to exist in local and global markets[1]. Even in Turkey, there will be more than 1 million companies registered by 2015. In such a competitive and harsh economic model, one of the problems that companies are challenged with is to protect their trademarks, in addition to the protection of their patents and inventions.

Although trademark retrieval is a very important problem for companies and it bears many challenges of Computer Vision and Pattern Recognition, existing datasets and methods are far from being sufficient. In the literature, the largest dataset for trademark retrieval includes 30,000 trademarks [17] and it is not publicly available. The publicly available datasets, on the other hand, include on the order of thousands of trademarks only - see Table 1. We believe that the non-availability of a large, challenging dataset had a negative influence on the pace of progress in trademark retrieval.

[1] See http://www.econstats.com/wdi/wdiv_494.htm for related statistics.

As stated by Kesidis et al. [6], trademark retrieval involves many challenges of different complexities and the problem is largely unsolved. Kesidis et al. suggested that an approach for trademark retrieval should address not only matching of low-level features extracted at keypoints but also recognition of text and more importantly, the ability to perform perceptual interpretations of logos, involving principles like Gestalt grouping laws. However, the existing datasets in the literature are far from evaluating such challenges.

In this article, we introduce a new, challenging dataset composed of more than 930,000 trademarks belonging to real companies. The dataset includes the challenges as discussed by Kesidis et al. [6], and we believe that, by benchmarking the trademark retrieval studies in this dataset, the researchers in the field will get the chance to develop improved methods and to achieve better performances. Moreover, we test the performances of the state of the art methods on the dataset and show that the best available method achieves around 60,000 mean average rank, which is farm from being practical.

1.1 Trademark Retrieval Studies

Like similar Computer Vision and Pattern Recognition problems, trademark retrieval (TR) studies can be broadly decomposed into two stages: feature extraction and matching. For the feature extraction phase, many features have been used: spatial distributions and frequencies of pixels [5], Fourier Descriptors [8], Zernike moments [20], curvature, centroid distance [20], shape context [16], scale-invariant feature transform (SIFT) [4, 7], orientation histogram [2] etc.

For the matching phase, the vectorized features are usually matched using distance metrics such as Euclidean, intersection, or using more complex matching methods such as deformable template matching [2].

2 A New Challenging Dataset

Our main trademark dataset includes 930,328 images, corresponding to 409,834 many different company trademarks .As illustrated in Figure 1, the dataset includes trademarks of different companies that not only include colored shapes but also text of various forms. The details of the dataset are described in Table 2, and the dataset is publicly available for research purposes at the following URL: http://kovan.ceng.metu.edu.tr/LogoDataset/

From the dataset, we have selected 320 similar trademark sets (similar to those in Figures 1b and 1c). From these, we see that it is far from trivial to find similarities by using local features only and that a good approach should combine color, texture, text, shape and parts information as much as possible, in addition

Table 1: Comparison of existing trademark datasets.

Dataset	Dataset Type	Number of logos (and images)	Types of images	Year	Publicly Available?
University of Maryland (UMD) [12]	Logo-to-logo	106 (-)	Bi-color	2001	Yes
BelgaLogos [3]	Logo-to-image	26 (10,000)	RGB	2009	Yes
Wei et al. [20]	Logo-to-logo	14 (1003)	Bi-color	2009	No
Flickr Logos [15]	Logo-to-image	32 (8,240)	RGB	2011	Yes
MICC Logos [18]	Logo-to-image	13 (720)	RGB	2013	Yes
EURO 2008 [7]	Logo-to-image	18 (106)	RGB	2010	No
METU Dataset	**Logo-to-logo**	**409,834 (930,328)**	**RGB**	**2014**	**Yes**

Table 2: Details of our dataset.

Aspect	Value
# trademarks	930,328
# unique registered firms	409,834
# unique trademarks	691,149
# trademarks containing text only	589,562
# trademarks containing figure only	19,394
# trademarks containing figure and text	312,154
# other trademarks	8,942

to the global information, to be able to achieve good performance.

3 Evaluated Feature Descriptors for Trademark Retrieval

In this section, we briefly explain the widely-used trademark-retrieval features that we evaluated on our dataset. For the sake of space, and since these methods are very well established in the Computer Vision and Pattern Recognition literature, we will skip the details as much as possible.

3.1 Color Histogram

Color is one of the widely-used source of information used in image retrieval [1, 9, 11, 14, 17]. We implemented 36 bits HSV color spatial histogram detailed in [9] with histogram intersection distance:

$$d^2\left(h,g\right) = 1 - \frac{\sum_i^n \min(h_i, g_i)}{\min(\left|h\right|, \left|g\right|)}. \qquad (1)$$

Compared to RGB color space, the HSV color space is closer to human vision, especially when quantized. Experimentally, we determined that 36 bits is ideal; more bits would increase computational cost, while less bits would decrease distinctiveness.

3.2 Gradient Orientation Histogram

Intensity gradients are also very informative of trademarks as shown by [1, 14]. For this, we first applied canny edge detector to the median-filtered image and calculate orientation of each gradient point. From the quantized orientation of each gradient point, we create a gradient orientation histogram. The quantization level of gradient point is essential, because too fine quantization is very sensitive to rotation, and too coarse quantization leads to a lack of distinctiveness. For comparing two gradient orientation histograms, we used Euclidean distance.

3.3 Local Binary Patterns

Most trademarks include textural information which can be best captured by texture descriptors. For this reason, we use Local Binary Patterns (LBP) [13], a successful and widely-used texture descriptor in the literature. We implemented 8-neighbor, rotation invariant, uniform LBP [13]. Normally, 8-neighbor LBP has 256 different patterns, while only 10 special patterns are rotation-invariant and uniform. For comparing the LBP vectors, we use Cosine vector distance.

3.4 Shape Context

Since shape is also important for trademarks, we employ a widely used feature for that: Shape context [16, 17]. Shape context uses N discrete sample points from internal and external contours of the shape for representing the shape. The shape context of a sample point is a vector of the relative positions of the other N-1 points. Since the original version of shape context is computationally costly, we take each context vector as a visual word and create the bag-of-visual-words (BOW) version of it [19].

We sampled 200 points from internal and external boundary of each trademark image, and for each point's shape context, we use five bins for the log-distance and 12 bins for relative orientation, which leads to 60 bit vectors for each point. These are then compared using Euclidean distance.

3.5 SIFT and Triangular SIFT

SIFT (scale-invariant feature transform) [10] is a well-known scale-invariant local feature descriptor, demonstrating better performance than many other feature descriptors in the literature. Since it is computationally heavy for matching, SIFT features are usually quantized into visual words [19].

In [4], triplets of SIFT features were used for trademark retrieval in order to incorporate local parts information into features. In this method, SIFT features at the same scale are grouped into triplets by using multi-scale Delaunay triangulation, and only triplets of SIFT features having the same scale are compared for finding similarities.

4 Experiments and Results

In this section, we evaluate the methods described in Section 3 on our dataset. We use precision-recall (PR), average rank and normalized rank (similar to

(a) Dataset samples

(b) Example set for similar trademarks

(c) Another example set for similar trademarks

Figure 1: Samples from our dataset. (a) Arbitrary samples. (b) Example for similar trademarks. (c) Another example set for similar trademarks.

[19]). Average rank of a retrieval is defined as follows:

$$\widetilde{Rank} = \frac{1}{N \times N_{rel}} \left(\sum_{i=1}^{N_{rel}} R_i - \frac{N_{rel}(N_{rel}+1)}{2} \right), \quad (2)$$

where N_{rel} is the number of relevant images for particular query image, N is the size of the image set, and R_i is the rank of the i^{th} relevant image. The normalized version of average rank is also useful for analyzing the performance of retrieval systems [19]:

$$Rank = \frac{1}{N_{rel}} \sum_{i=1}^{N_{rel}} R_i. \quad (3)$$

Average rank measures takes values in the range from $1 + \frac{N_{rel}}{2}$ to $N - \frac{N_{rel}}{2}$. In contrast, the normalized rank measure lies in the range $[0, 1]$, such that being close to 0.0 means good retrieval, and above 0.5 corresponds to random retrieval.

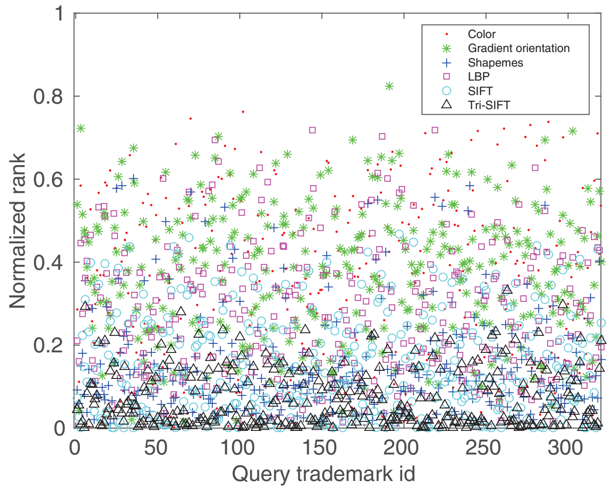

Figure 2: Normalized ranking result of the performance of the method on our dataset.

4.1 Performance Comparison

Figure 2 plots the normalized rank values of the similar trademarks found by the different methods. We see that SIFT and Tri-SIFT perform better than other methods, and color and gradient orientation histogram do not perform well.

When we look at the precision-recall values in Figure 3, we see a similar trend and more clearly that SIFT performs better than Tri-SIFT in terms of precision-recall values. However, when we analyze the values in Table 3 for the performances of methods measured by average rank and normalized rank, we see that Tri-SIFT yields better results. The reason is that SIFT can find the desired logos in top 1% most of the time but when it makes a mistake, it makes a huge one, finding similarities in the bottom 1%. Tri-SIFT, on the other hand, does not make as worse mistakes as SIFT; however, it cannot find similarities as good as SIFT. For this reason, precision-recall and rank measures yield different values for SIFT and Tri-SIFT.

Another important aspect of the methods is their running time. For this, we run a test on pre-computed features for the trademarks in the dataset, and analyze only the matching performance. When we look at how much each method takes for comparing a query to each trademark in the dataset (listed in Table 4), we see that, as expected, TRI-SIFT is the slowest whereas simpler features like color histogram, gradient orientation histogram and LBP are very fast compared to SIFT, TRI-SIFT and shapemes.

Table 3: Comparison of the methods on our dataset.

Algorithm	Average rank	Normalized rank
Color	314,953.2 ± 194,291.3	0.339 ± 0.209
Orientation	350,662.1 ± 149,797.4	0.377 ± 0.161
LBP	244,830.3 ± 133,210.8	0.263 ± 0.143
SHAPEME	141,489.5 ± 117,323.3	0.152 ± 0.126
SIFT	141,994.2 ± 116,035.8	0.153 ± 0.125
TRI-SIFT	**66,117.9** ± 64,736.4	**0.071** ± 0.070

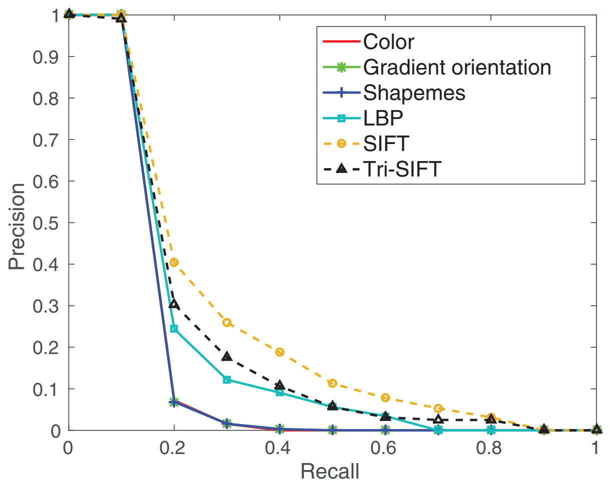

Figure 3: Precision-recall of the performance of the methods on our dataset. [Best viewed in color]

Table 4: Running time of the methods on our dataset. The table lists only the matching time of a query to 930,328 logos in the dataset (in Matlab).

Algorithm	Time (milliseconds)	Parallel process
Color	141.6	No
Orientation	100.3	No
LBP	49.4	No
SHAPEME	18,567.7	Yes
SIFT	19,195.9	Yes
TRI-SIFT	53,292.8	Yes

5 Conclusion

In this article, we introduced a new big challenging dataset for trademark retrieval composed of 930,328 trademark images. Using the state-of-the-art features used in the literature, we showed that the dataset is very challenging and that it can be used for benchmarking the trademark retrieval studies.

Using different measures, our comparisons showed that TRI-SIFT and SIFT perform best among the methods tested in the article. However, compared to other features, they are slower; nonetheless, this seems to be a minor issue since SIFT and its extension Tri-SIFT take less than 1 minute when a query is compared against all the trademarks in the dataset.

6 Acknowledgments

This work is supported by the Ministry of Science, Turkey, under the project SANTEZ-0029.STZ.2013-1. We would like to thank Usta Bilgi Sistemleri A.Ş. and Grup Ofis Marka Patent A.Ş. for their contributions to the project, by especially making the dataset available for us and the whole community.

References

[1] A. K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, Aug. 1996.

[2] A. K. Jain and A. Vailaya. Shape-Based Retrieval: a Case Study With Trademark Image Databases. *Pattern Recognition*, 31(9):1369–1390, Sept. 1998.

[3] A. Joly and O. Buisson. Logo retrieval with a contrario visual query expansion. In *17th ACM international conference on Multimedia*, pages 581–584, 2009.

[4] Y. Kalantidis, M. Trevisiol, L. G. Pueyo, R. van Zwol, and Y. Avrithis. Scalable triangulation-based logo recognition. *1st ACM Int. Conference on Multimedia Retrieval - ICMR*, pages 1–7, 2011.

[5] T. Kato. Database architecture for content-based image retrieval ages with each user. *Int. Symposium on Electronic Imaging, Science and Technology*, pages 112–123, 1992.

[6] A. Kesidis and D. Karatzas. *Handbook of Document Image Processing and Recognition.* Springer London, London, 2014.

[7] M. Köstinger, P. M. Roth, and H. Bischof. Planar trademark and logo retrieval. *TR, ICGTR10/01, Graz*, 2010.

[8] C. P. Lam, J. K. Wu, and B. Mehtre. *STAR a system for trademark archival and retrieval.* ACCV, 1995.

[9] Z. Lei, L. Fuzong, and Z. Bo. A CBIR method based on color-spatial feature. *Multimedia Technology for Asia-Pacific Information Infrastructure*, 1:166–169, 1999.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[11] C. W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. Qbic project: querying images by content, using color, texture, and shape. In *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, 1993.

[12] U. of Maryland Logo Dataset. Laboratory for language and media processing (lamp), online: `http://lampsrv02.umiacs.umd.edu/projdb/project.php?id=47`, Last accessed: 18 December, 2014.

[13] T. Ojala, M. Pietikäinen, and T. Mäenpää. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In *ICPR*. Springer, 2001.

[14] R. Phan and D. Androutsos. Content-based retrieval of logo and trademarks in unconstrained color image databases using Color Edge Gradient Co-occurrence Histograms. *CVIU*, 114(1):66–84, jan 2010.

[15] S. Romberg, L. G. Pueyo, R. Lienhart, and R. van Zwol. Scalable logo recognition in real-world images. In *ACM ICMR*, 2011.

[16] M. Rusiñol and J. Lladós. Efficient logo retrieval through hashing shape context descriptors. *8th IAPR Int. Workshop on Document Analysis Systems*, 2010.

[17] M. Rusinol, D. Aldavert, D. Karatzas, R. Toledo, and J. Lladós. Interactive trademark image retrieval by fusing semantic and visual content. In *Advances in Information Retrieval*, pages 314–325. Springer, 2011.

[18] H. Sahbi, L. Ballan, G. Serra, and A. Del Bimbo. Context-dependent logo matching and recognition. *IEEE Transactions on Image Processing*, 22(3):1018–1031, Mar. 2013.

[19] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. *ICCV*, 2003.

[20] C. H. Wei, Y. Li, W. Y. Chau, and C. T. Li. Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, 42(3):386–394, Mar. 2009.