Interactive Trademark Image Retrieval by Fusing Semantic and Visual Content

Marçal Rusiñol, David Aldavert, Dimosthenis Karatzas, Ricardo Toledo and Josep Lladós

Computer Vision Center, Dept. Ciències de la Computació
Edifici O, Univ. Autònoma de Barcelona
08193 Bellaterra (Barcelona), Spain
{marcal,aldavert,dimos,ricard,josep}@cvc.uab.cat

Abstract. In this paper we propose an efficient queried-by-example retrieval system which is able to retrieve trademark images by similarity from patent and trademark offices' digital libraries. Logo images are described by both their semantic content, by means of the Vienna codes, and their visual contents, by using shape and color as visual cues. The trademark descriptors are then indexed by a locality-sensitive hashing data structure aiming to perform approximate k-NN search in high dimensional spaces in sub-linear time. The resulting ranked lists are combined by using a weighted Condorcet method and a relevance feedback step helps to iteratively revise the query and refine the obtained results. The experiments demonstrate the effectiveness and efficiency of this system on a realistic and large dataset.

Keywords: Multimedia Information Retrieval, Trademark Image Retrieval, Graphics Recognition, Feature Indexing.

1 Introduction

The digital libraries of patent and trademark offices contain millions of trademark images registered over the years. When registering a new trademark it is important to browse these databases in order to be sure that there is no other company having a similar logo design so as to avoid infringing someone else's intellectual property rights. However, nowadays the means we have to browse and retrieve information from these databases are not really useful enough to assess if a logo to be registered might provoke trademark infringement. Usually, the way that Intellectual Property Offices offer to navigate through the trademark image collections is by the use of subject terms that aim to catalogue the image contents. Each logo is labelled with a set of predefined metadata which enable to search in the collection by the selection of a set of labels. The most widely used metadata classification codes in the Intellectual Property Offices are the ones from the Vienna classification system, developed by the World Intellectual Property Organization [16]. This manually annotation of the logos' contents has the advantage of clustering the logos by semantic information. This categorization system is hierarchical so the category definition can be specified from broad concepts to more specific contents. As an example in the broad categories we can find *Category 1: Celestial bodies, natural phenomena, geographical maps*, and under this category we find for instance *class 1.7.6: Crescent moon, half-moon* or *class 1.15.3: Lightning*. Another example for broad categories would be *Category 17: Horological instruments, jewelry, weights and measures*, where on the more specific classes we find, *class 17.2.13: Necklaces, bracelets, jewelry chains*, or *class 17.1.9: Clocks, alarm clocks*. However this classification approach presents some limitations, specially for abstract or artistic images, where the use of manual classification codes is not always distinctive. The user that browses a trademark image database is usually looking for images that *look* similar to the one that is willing to register, but images in the same semantic category are usually far from being similar one to each other, so a tedious manual inspection of all the logos in a given category is still needed to retrieve near-duplicate images. In that specific scenario it would be very interesting if we could enhance this existing retrieval mechanism with techniques from the Content-based Image Retrieval (CBIR) domain which help us to describe trademark images in terms of their visual similarity to the given query.

Since the early years of CBIR, a lot of effort has been devoted to the particular application of trademark image retrieval. Systems like STAR [17], ARTISAN [3] or TAST [11] were proposed to be used in different trademark offices to index logo images by visual similarity. These systems just focused on binary device mark images and use shape and texture as the only indicators of similarity. In the latest years some attempts to merge shape and color for the retrieval of trademark images have been proposed. For instance, the method presented by Hsieh and Fan in [8] use the RGB color values as input for a region growing algorithm that further characterizes the trademark designs with a shape description of these color regions. In [6], Hitam et al. propose to use the spatial distribution of RGB values to describe trademarks by means of which colors compose them and where they appear. More recently, Hesson and Androutsos proposed in [5] to describe trademark images with the use of a color naming algorithm instead of the raw RGB values in order to provide a response that agrees more with the human perception of colors. From the perceptual point of view, works like [14], [7] or [9] introduce the use of the Gestalt theory to the description of trademarks. All these systems perform well at retrieving relevant images by visual similarity, but they all discard semantic information that in some cases might also be very relevant. To our best knowledge very few attempts have been made to combine semantic and visual information in a trademark retrieval system. Ravela et al. presented in [13] a system that retrieves binary trademark images by restricting the search to a set of predefined semantic categories. In our paper we extend this idea by proposing a method that takes into account semantic information without restricting the search corpus to a specific subset of images. In addition, relevance feedback is used to refine the semantic and visual cues.

The main contribution of this paper is to present an efficient queried-by-example retrieval system which is able to retrieve logos by visual similarity and semantic content from large databases of isolated trademark images. Logos are

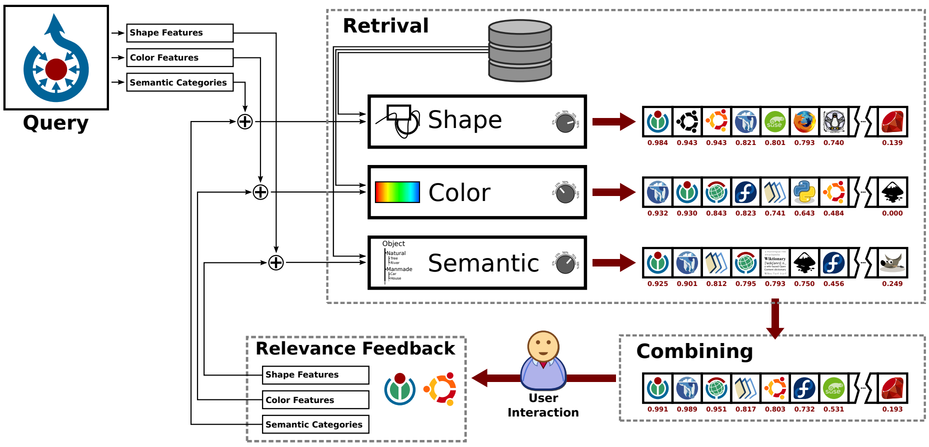


Fig. 1. Overview of the proposed system

compactly described by descriptors of their shape and color. The semantic description of the trademark images is given by a hierarchical organization of the Vienna codes. These descriptors are then organized by a locality-sensitive hashing indexing structure aiming to perform approximate k-NN search in high dimensional spaces in sub-linear time. The use of a hashing technique allow us to quickly index and retrieve queried logos by visual similarity and semantic content. The resulting ranked lists are then fused by using the Condorcet method and a relevance feedback step helps to iteratively revise the query and refine the obtained results. To conduct the experimental results, we will focus on a large collection of real-world trademark images registered to the Spanish Intellectual Property Office.

The remainder of this paper is organized as follows: section 2 is devoted to present an overview of the system. In section 3 the logo description using visual cues is presented, while in section 4 the semantic representation provided by the Vienna codes is introduced. Subsequently, in section 5, the indexing structure for efficient retrieval, the ranking combination algorithm and the relevance feedback strategy are presented. Section 6 provides the experimental results and finally section 7 is a summary and discussion of extensions and future work.

2 System Overview

We can see in Figure 1 an overview of the proposed system. The user provides a query trademark image he wants to register and, optionally, a set of semantic concepts from the Vienna classification codes. Shape and color features are extracted from the query image and an indexing mechanism efficiently retrieves from the database the trademarks that have a shape or color description similar to the query one. These ranked lists are combined to form a single resulting list which is presented to the user. The user can then refine the search by selecting the logo images that are considered to be relevant. This relevance feedback mechanism allows the user's search query to be revised in order to include a percentage of relevant and non-relevant documents as a means of increasing the search engine's precision and recall. The relevance feedback step is also a way to weight the importance of a visual cue. In some cases the color might be relevant to the user despite the shape (dis)similarity or viceversa. In addition, in the relevance feedback step, if the user did not provide any semantic concept in the query, these are deduced from the selection of relevant documents and used in the following iterations in order to obtain a refined result list. If the user starts selecting trademarks from a given semantic class despite being dissimilar in both shape and color, the system automatically adapts the queries and the combination of the results to the user needs, giving more importance to the semantics than to the visual information. After this brief overview, let us present each of the system modules, starting with the visual description of trademark images.

3 Visual Description of Trademark Images

We base our visual description of logo images on two separate descriptors, one encoding shape information and the other describing the colors of the trademark designs. For the shape information we use the shape context descriptor and for the color we use a quantization of the $CIEL^*C^*h$ color space. Let us first briefly overview the shape context descriptor, and then focus on how we describe trademark images by color.

3.1 Shape Information

The shape context descriptor was proposed by Belongie et al. in [1]. This descriptor allows to measure shape similarity by recovering point correspondences between the two shapes under analysis. In a first step, a set of interest points has to be selected from the logos. Usually, a Canny edge detector is used and the edge elements are sub-sampled in order to obtain a fixed number of n points p_i per logo ℓ. For each of these n points, the shape context captures the edge point distribution within its neighborhood. A histogram using log-polar coordinates counts the number of points inside each bin. For a point p_i of the shape, a histogram h_i of the coordinates of the nearby points q is computed as:

$$h_i(k) = \#\{q \neq p_i : q \in bin_{p_i}(k)\} \tag{1}$$

In our experimental setup, we have chosen 5 bins for $\log r$ and 12 bins for θ. The descriptor offers a compact representation of the distribution of points relative to each selected point. Once all the n points in a shape are described by their shape context descriptor, we compute the shapeme histogram descriptor presented by Mori et al. in [12] to efficiently match two trademarks. This description technique was inspired by the shape context descriptor described above, and the bag-of-words model. The main idea is to compute the shape context descriptor for all the interest points extracted from a symbol and then use vector quantization in the space of shape contexts. Vector quantization involves a clustering stage of the shape context feature space. Once the clustering is computed, each shape context descriptor can be identified by the index of the cluster which it belongs to. These clusters are called shapemes. Each logo is then described by a single histogram representing the frequencies of appearance of each shapeme.

In the learning stage, given a set of model logos, we compute their shape context descriptors and cluster this space by means of the k-means algorithm, identifying a set of k cluster centers and assigning to them a given integer index $I \in [1, k]$. Then, in the recognition stage, given a logo ℓ, and its n sampled points from its edge map, we compute their shape context descriptors $h_i, \forall i \in [0, n]$. Each shape context descriptor of the points p_i is then projected to the clustered space and can be identified by a single index I_i. The logo ℓ can thus be represented by a single histogram coding the frequency of appearance of each of the k shapeme indices. By this means, we globally describe by a unique histogram SH each logo by applying the following equation:

$$SH(x) = \# \{I_i == x : I_i \in [0, k]\} \quad (2)$$

By using the shapeme histogram descriptor, the matching of two logos is reduced to find the k-NN in the space of shapeme histograms, avoiding much more complex matching strategies. We can see an example in Table 1 of the performance of this descriptor to retrieve trademark images by shape.

3.2 Color Information

In order to represent the color information of the trademark images, we use a color quantization process in a color space that is perceptually uniform, i.e. distances in this space agree with the human perception on whether two colors are similar or not.

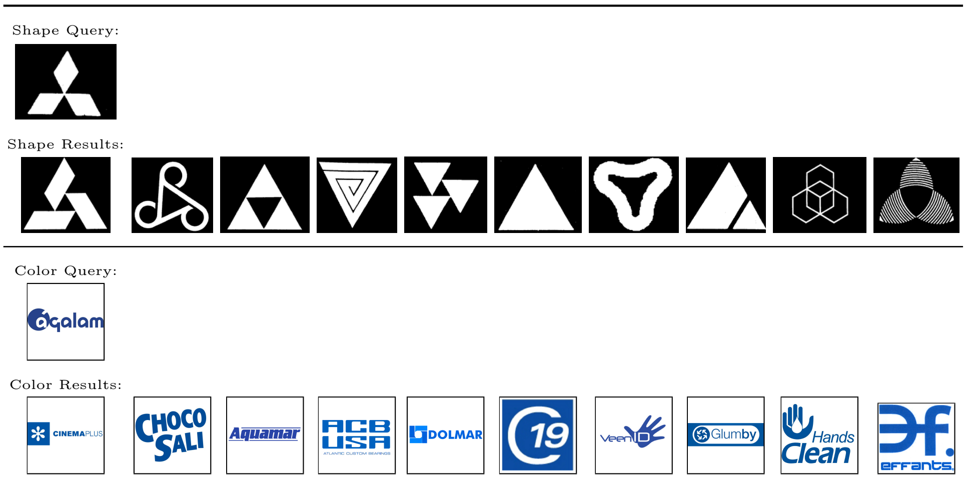
We transform the RGB trademark image to the $CIEL^*C^*h$ color space assuming D65 as a reference illuminant. Details on the formulae can be found in [18]. The L^* component corresponds to the lightness, the C^* component to the chroma and the h component to the hue. After having converted the image to a perceptually uniform color space, we compute the color descriptor of the image as a histogram of appearances of different colors. As we did with the shape descriptor, the k-means method clusters the $CIEL^*C^*h$ color space. Once the clustering is computed, each pixel value can be identified and quantized by the index of the cluster which it belongs to. Each logo is then described by a single histogram representing the frequencies of appearance of each color representatives.

In the learning stage, given a set of model logos, we convert the images to the $CIEL^*C^*h$ color space and cluster this space by using the k-means algorithm, identifying a set of q color representatives with indices $J \in [1, q]$. Then, in the recognition stage, given a logo ℓ converted to the $CIEL^*C^*h$ color space, each pixel P_{xy} is projected to the clustered space and it is identified by an index J_{xy}. The logo ℓ is represented by a single histogram coding the frequency of appearance of each of the q color indices. We globally describe by a unique histogram CO each logo by applying the following equation:

$$CO(x) = \#\left\{J_{xy} == x : J_{xy} \in [0, q]\right\} \tag{3}$$

By using the color histogram descriptor, the matching of two logos is reduced to find the k-NN in the space of color histogram. We can see an example in Table 1 of the performance of this descriptor to retrieve trademark images by color.

Table 1. Queries and first ten results by shape and color similarity respectively.



4 Description of Trademark Images by Vienna Codes

The Vienna classification codes [16] offer a hierarchical categorization of the image contents. In our system, all the trademarks in the database have an associated list of Vienna codes. We can see in Figure 2 a set of trademark images all belonging to the same category. As we can appreciate, there are few visual similarities yet all contain a horse. For a given trademark image and its associated Vienna code, we will define four different sets $S_1 ... S_4$ that cluster the dataset depending on its semantical agreement. In S_1 we will have all the trademark images under the same Vienna category, for the trademarks in the example that would be all the trademarks in category 03.03.01 (containing horses or mules). In S_2 we store all the trademark images that are under category 03.03 (horses, mules, donkeys, zebras) without being in S_1, in S_3 all the trademarks under category 03 (Animals). Finally, S_4 is the set of trademark images without any semantic agreement with the query.

Fig. 2. Example of searching logos by using the category tree of Vienna codes (a); and example of all the associated Vienna codes for a given trademark image (b).

Since a trademark image can have more than one Vienna code, each trademark image is stored in the set that agrees more with the query. Note that for the semantic categorization of trademarks, the retrieval result does not produce a ranked list, but sets of images. The only notion of order here is that elements in S_1 are better ranked than elements in S_2 and so on.

5 Retrieval Stage

In this section, we are going to present in detail the retrieval stage of the proposed method. We first introduce the indexing mechanism that allows to retrieve trademarks by visual similarity. Then, we present how visual and semantic cues are combined together. Finally, we detail the relevance feedback step where the user interaction is taken into account in order to refine the obtained results.

5.1 Indexing Trademark Visual Descriptors with LSH

In order to avoid a brute force matching of the shape and color feature vectors, we propose to use an algorithm for approximate k-NN search. This allows to efficiently obtain a set of candidates that probably lie nearby the queried point. The method we use is the locality-sensitive hashing (LSH), first introduced by Indyk and Motwani in [10], and then revised by Gionis et al. in [4]. The LSH algorithm has been proven to perform approximate k-NN search in sub-linear time.

The basic idea of the LSH method is to index points from a database by using a number l of naïve hash functions g, in order to ensure that objects close in the feature space have a high probability of provoking collisions in the hash tables. Given a query feature vector q the algorithm iterates over the l hash functions g. For each g considered, it retrieves the data points that are hashed into the same bucket as q. In our experimental setup, we are using $l = 50$ simple hash functions. We can see in Figure 3 a comparison of the average time taken to retrieve the 50 topmost similar vectors for different sizes of the database.

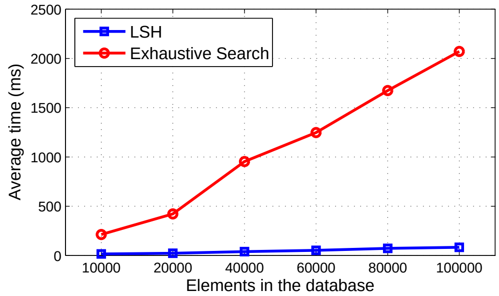


Fig. 3. Average time to retrieve the 50-NN for different database sizes.

5.2 Ranking Combination and Relevance Feedback

The retrieval of trademark images either by shape or by color result in a ranked list. Those lists need to be combined to return to the user a single result list. However, the semantic part do not provide the results in a ranked list but in separate sets. In order to combine these results we choose to use a weighted Condorcet method [2]. The particularity of this method compared to other classic combination algorithms, is that in can handle tied results.

The Condorcet algorithm is a method which specifies that the winner of the election is the candidate that beats or ties with every other candidate in a pair-wise comparison. In our case, we consider all the trademarks in the same semantic set as tied whereas the trademarks in upper sets are beaten. Given a query trademark and the ranked lists for the shape, the color and the ordered sets for the semantic retrieval, to obtain a final rank of trademark images we use their win and lose values. If the number of wins that a logo has is higher than the other one, then that trademark wins. Otherwise if their win property is equal we consider their lose scores, the trademark image which has smaller lose score wins. By letting the user weight the contribution of color, shape and semantics, we obtain a weighted Condorcet method. The weighting factor allow to fuse the ranked lists taking into account the user's needs, i.e. selecting the information cue more relevant to its query. Albeit the computational cost of the Condorcet algorithm is high, we limit the comparisons of wins and looses to the best subset of elements returned by the LSH algorithm instead of using the full dataset corpus (in our experiments limited to best 100 logos by shape and by color). Therefore, the cost of applying the Condorcet algorithm is kept bounded.

After applying the Condorcet algorithm, the user receives a single sorted list with the trademarks that are closer to the query in terms of visual similarity (shape and color) and semantic content. Once we present this list to the user, we include a relevance feedback step so as to give the chance to the user to refine the results.

Relevance feedback should be a must for any trademark image retrieval application. In our case, in order to include the feedback from the user, we use Rocchio's algorithm [15] to revise the vector queries and weight the importance of shape, color and semantics depending on the user needs. At each iteration the Rocchio's algorithm computes a new point in the query space aiming to incorporate relevance feedback information into the vector space model. The modified query vector $\overrightarrow{Q_m}$ is computed as

$$\overrightarrow{Q_m} = \left(\alpha * \overrightarrow{Q_o}\right) + \left(\beta * \frac{1}{|D_r|} * \sum_{\overrightarrow{D_j} \in D_r} \overrightarrow{D_j}\right) - \left(\gamma * \frac{1}{|D_{nr}|} * \sum_{\overrightarrow{D_k} \in D_{nr}} \overrightarrow{D_k}\right) \quad (4)$$

where $\overrightarrow{Q_o}$ is the original query vector, and D_r and D_{nr} the sets of relevant and non-relevant trademark images respectively. α, β and γ are the associated weights that shape the modified query vector respect the original query, the relevant and non-relevant items. In our experimental setup we have chosen the following values $\alpha = 0.55$, $\beta = 0.4$ and $\gamma = 0.05$ in order to keep $\overrightarrow{Q_m}$ normalized and within the feature space.

6 Experimental Results

Let us first introduce the logo dataset we used in our experiments. We will then present the experimental results.

6.1 Trademark Dataset

To conduct the experimental results, we will focus on a collection of real trademark images downloaded from the Spanish Intellectual Property Office[1] with their associated Vienna codes. This dataset is composed by all the trademarks registered during the year 2009, that is near 30000 trademark images organized within 1350 different Vienna categories. In average, each trademark image has associated 2.17 Vienna codes. Another subset of 3000 trademark images has been used as training set to run the k-means clustering algorithm to build both the shape and color descriptors.

6.2 Method Evaluation

The first experiment presents in Figure 4 a qualitative evaluation of the proposed method. For the first query, we search for near-duplicate images in the database. In that case, the visual cues are the most important despite some trademarks from the semantic category "star" appear. Surrounded in green we present the trademarks selected as relevant by the user at each iteration. As we can see, the results are qualitatively better iteration after iteration. In the second query we show in Figure 4, we do not expect to retrieve near-duplicates but perceptually similar logo images. The user selects trademarks which contain a tree with some text somewhere in the image. In that case, initially the proposed method fails

[1] http://www.oepm.es

Fig. 4. Retrieval results obtained looking for near-duplicate trademarks in the first query and semantic similar logos in the second.

in the topmost images, but iteration after iteration the results are correctly reranked.

The second experiment provides a simple quantitative evaluation of the use of the relevance feedback step. We selected six different trademarks that have several (more than 10) near-duplicate entries in the database as our queries. These near-duplicate images are labelled as being the only relevant answers that the system might provide. With this groundtruth data, we are able to compute the precision and recall measures. We can see in Figure 5 the evolution of the mean average precision and recall measures (computed at the topmost 30

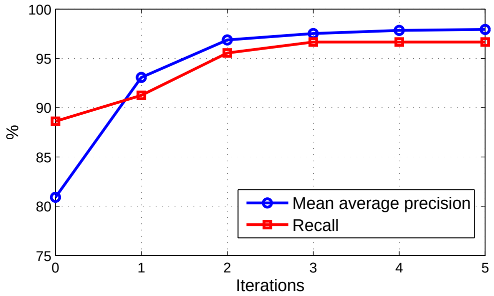


Fig. 5. Mean average precision and mean recall at different iterations.

elements in return) through the successive iterations of the loop composed by retrieval and further user feedback. As we can appreciate, both measures present an important increase before reaching stability.

7 Conclusions

In this paper, we have presented an interactive trademark image retrieval system which combines both visual and semantic information to obtain the most similar logos from a large realistic database. The addition of the semantic information provided by the standard Vienna codes increases the retrieval performance, by reducing the amount of possible similar logos and also enabling the method to retrieve trademarks which are dissimilar in shape or color content but have an strong semantic connection. Besides, the addition of user feedback allows to further refine the obtained results and it permits to automatically generate the semantic descriptor of the query when it is not given by the user. The qualitative evaluation of the method show a good performance of the system in retrieving trademarks by combining visual terms and semantic concepts. Besides, the use of relevance feedback steadily increases the mean average precision and recall of the system when used for searching near-duplicate trademark images in the database.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Education and Science under projects TIN2008-04998, TIN2009-14633-C03-03, TRA2010-21371-C03-01, Consolider Ingenio 2010: MIPRCV (CSD200700018) and the grant 2009-SGR-1434 of the Generalitat de Catalunya.

References

1. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.

2. M. de Condorcet. *Essai sur l'application de l'analyse à la probabilité des decisions rendues à la pluralité des voix.* 1785.
3. J.P. Eakins, K. Shields, and J. Boardman. ARTISAN: A shape retrieval system based on boundary family indexing. In *Storage and Retrieval for Still Image and Video Databases IV*, volume 2670 of *Proc. SPIE*, pages 17–28. 1996.
4. A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimensions via hashing. In *Proceedings of the 25th International Conference on Very Large Data Bases*, pages 518–529, 1999.
5. A. Hesson and D. Androustos. Logo and trademark detection in images using color wavelet co-occurrence histograms. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1233–1236, 2008.
6. M.S. Hitam, H.J.W. Yussof W.N. Jawahir, and M.M. Deris. Hybrid Zernike moments and color-spatial technique for content-based trademark retrieval. In *Proceedings of the International Symposium on Management Engineering*, 2006.
7. V.J. Hodge, G. Hollier, J. Austin, and J. Eakins. Identifying perceptual structures in trademark images. In *Proceedings 5th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications*, 2008.
8. I.S. Hsieh and K.C. Fan. Multiple classifiers for color flag and trademark image retrieval. *IEEE Trans. Image Process.*, 10(6):938–950, 2001.
9. H. Iguchi, K. Abe, T. Misawa, H. Kimura, and Y. Daido. Recognition of grouping patterns in trademarks based on the Gestalt psychology. *Electron. Commun. Jpn.*, 92(10):49–60, 2009.
10. P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 604–613, 1998.
11. P.W.H. Kwan, K. Toraichi, K. Kameyama, F. Kawazoe, and K. Nakamura. TAST-trademark application assistant. In *Proceedings of the International Conference on Image Processing*, pages 884–887, 2002.
12. G. Mori, S. Belongie, and J. Malik. Efficient shape matching using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(11):1832–1837, 2005.
13. S. Ravela and R. Manmatha. Multi-modal retrieval of trademark images using global similarity. Technical Report UM-CS-1999-032, University of Massachusetts, 1999.
14. M. Ren, J.P. Eakins, and P. Briggs. Human perception of trademark images: Implications for retrieval system design. *J. Electron. Imaging*, 9(4):564–575, 2000.
15. J. Rocchio. Relevance feedback in information retrieval. In *SMART Retrieval System: Experiments in Automatic Document Processing*, pages 313–323, 1971.
16. J. Schietse, J.P. Eakins, and R.C. Veltkamp. Practice and challenges in trademark image retrieval. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, pages 518–524, 2007.
17. J.K. Wu, B.M. Mehtre, Y.J. Gao, P.C. Lam, and A.D. Narasimhalu. STAR - a multimedia database system for trademark registration. In *Applications of Databases*, volume 819 of *Lecture Notes on Computer Science*, pages 109–122. 1994.
18. G. Wyszecki and W.S. Stiles. *Color science: Concepts and methods, quantitative data and formulae.* JohnWiley and Sons, New York, 1982.