Logo and Trademark Recognition

18

Anastasios Kesidis and Dimosthenis Karatzas

Contents

A. Kesidis (✉)
Department of Surveying Engineering, Technological Educational Institution of Athens, Aigaleo, Athens, Greece
e-mail: akesidis@teiath.gr

D. Karatzas
Computer Vision Center, Universitat Autònoma de Barcelona, Bellaterra, Spain
e-mail: dimos@cvc.uab.es

D. Doermann, K. Tombre (eds.), *Handbook of Document Image Processing and Recognition*, DOI 10.1007/978-0-85729-859-1_16,

Abstract

The importance of logos and trademarks in nowadays society is indisputable, variably seen under a positive light as a valuable service for consumers or a negative one as a catalyst of ever-increasing consumerism. This chapter discusses the technical approaches for enabling machines to work with logos, looking into the latest methodologies for logo detection, localization, representation, recognition, retrieval, and spotting in a variety of media. This analysis is presented in the context of three different applications covering the complete depth and breadth of state of the art techniques. These are trademark retrieval systems, logo recognition in document images, and logo detection and removal in images and videos. This chapter, due to the very nature of logos and trademarks, brings together various facets of document image analysis spanning graphical and textual content, while it links document image analysis to other computer vision domains, especially when it comes to the analysis of real-scene videos and images.

Keywords

Logo recognition • Logo removal • Logo spotting • Trademark registration • Trademark retrieval systems

Introduction

Logos and trademarks offer a particularly interesting setting for document analysis applications. By their very nature, they can incorporate any combination of textual and graphical content; subsequently, their treatment typically spans different areas of document analysis. At the same time, the range of applications centered on logo and trademark detection and recognition stems from various, quite different in nature necessities. In this chapter the main species of such applications are defined, and related methodologies and best practices are reviewed.

The word "trademark" is a legal concept that covers any type of indicator that can distinguish the products or services (in which case it is called a "service mark") of an entity from those of the competitors. Trademarks in the wider sense can in principle be defined over any modality such as sound, movement, a distinctive color, or individual words, as long as they are capable of distinguishing goods or services in a particular market segment.

A logo on the other hand is a graphic representation such as a symbol, badge, emblem, icon, or picture. Formally, when a logo comprises just an arranged typeface (e.g., the Microsoft logo), it is called a logotype, although habitually the term logo is used to denote any combination of graphic representation and arranged typeface. A logo can be used as a trademark if a business opts to register it as such.

The terms "logo," "trademark," "service mark," "certification mark," "brand," "mark," and "label" are often used interchangeably, although there are key differences as far as their legal definition is concerned. In this chapter the terms "logo" and "trademark" are used without making any special distinction between them, to refer to any combination of graphic representation and arranged typeface used to distinguish an entity, product, or service.

History and Importance

The use of distinctive marks to designate ownership goes a long way back in time with livestock ownership marks and pottery seals use dating back to the fourth millennium BC. After the fall of the Roman Empire, marks were increasingly used to identify product makers with a guild, which eventually led to the recognition of the property value of marks as a means to carry the reputation of a maker. The earliest enactment on marks was probably the Bakers Marking Law, introduced in England in 1266, which required bakers to mark their work with either pinpricks or stamps, followed by a similar requirement on silversmiths in 1363. Nevertheless, no strict legal framework protecting the trademark value was implemented until the early fifteenth and sixteenth centuries. In the USA, Thomas Jefferson advocated trademark laws in 1791, but it took until 1870 to pass the law on registration of trademarks and 1905 for the US Patent and Trademark office to be created. In the UK, the law on registration of trademarks was enacted in 1875. Numerous contenders exist for the earliest, continuously used mark, including Lowenbrau (since 1383) and Stella Artois (1366), while the oldest registered trademark in the world is considered to be the mark of "Bass & Co," registered in the UK in 1875.

Nowadays, national or regional Intellectual Property offices around the world deal with the registration of trademarks, while the legal framework has been revised to serve the needs of a globalized market. The Madrid system for the international registration of marks was established in 1891 and offers a trademark owner the possibility to have his trademark protected in several countries. The Madrid system is administered by the World Intellectual Property Organization (WIPO), a specialized agency of the United Nations. In the EU, the Office of Harmonization for the Internal Market (OHIM) was established in 1994 to deal with European Union-wide trademark registrations.

According to the World Intellectual Property Organization, trademark applications per year worldwide have been increasing almost linearly over the past 25 years reaching about 3.5 million applications and 1 million new trademark registrations in 2010. The total number of trademarks in force around the world is estimated to

Table 18.1 Trademark registration statistics for the year 2010 and selected IP offices. Note that new trademarks registered can exceed the number of applications due to back processing applications of previous years

IP office	Trademarks in force	Trademark applications	Trademark registrations
China	4,603,995	1,057,480	1,333,097
Japan	1,751,854	124,726	102,597
United States of America	1,544,184	281,867	167,641
France	1,119,000[a]	93,187	4,250
Spain	887,122	47,120	41,092
OHIM (European Union)	609,373	98,616	102,227
Australia	446,760	59,459	39,943
Russian Federation	392,202	56,856	40,136
Germany	N/A	74,339	53,300
Worldwide estimates (based on available data from 188 IP offices)	*20,259,654*	*3,481,763*	*2,954,227*

Source: "2011 World Intellectual Property Indicators," World Intellectual Property Organization Publication No. 941E/2011, ISBN 978-92-805-2152-8, 2011
[a]Data available for 2009 only

exceed 20 million (see Table 18.1). The vast majority of trademark registrations is graphical or combined textual and graphical.

Considering the above statistics, it is easy to appreciate the significant economic and social impact of trademark registration and use. Apart from the vast amount of man-months invested in managing the registration of new trademarks, a number of derivative services and applications exist that complement the trademark ecosystem. For example, enforcing the rights of ownership of registered trademarks is a commercial necessity that has resulted to the birth of trademark watch services, while numerous applications have been developed to leverage the recognizability of trademarks and logos to automate tasks such as incoming document categorization in digital mailroom implementations.

The automation of such services and applications is not trivial, as can be witnessed by the continuous investment in important research and development activities related to trademark recognition. Just to mention a few recent related projects, the EU project eMAGE (eContent programme, 1.1m €, 2004–2006) followed by eMARKS (eTEN programme, 2m €, 2007–2009) looked into a search service for images protected by IPR (trademarks and industrial designs). More recently, the EU project PROFI (FP6-IST, 1m €, 2004–2007) researched into perceptually relevant retrieval of figurative images (clip art, logos, signs). Over 2010–2012 the OHMI Cooperation Fund funded a 1.2m € project called

"Search Image" aiming to build a service to improve the current results of the figurative trademarks and design searches, revisiting among others the possibility to use computer vision-based techniques (until then considered "not mature enough").

Applications

The range of applications related to logos and trademarks stems from various, quite different in nature necessities. This section offers an overview of the main types of applications and the specific needs they address.

There are three main actors in the logo and trademark ecosystem: the trademark owners, the professionals that deal with the legal aspect of trademark management and registration, and a variety of end users that make use of logos to automate or facilitate daily tasks and processes.

The main concern of logo owners is to ensure the correct and efficient use of their logos. First, businesses want to make sure that their logos are seen by their target audience, which translates to achieving a certain degree of visibility and recognition. Second, they want to ensure that their corporate image is not misused, which implies monitoring and detecting any potentially wrong or illegal use of their protected marks. Typical applications required by logo owners relate to the detection of a small number of known a priori logos in unconstraint contexts like videos of events, or press images, but also trademark watch services where new trademark registrations are continuously monitored so that applications for similar trademarks can be identified and opposed to.

On the other end of the spectrum are the target audiences, the consumers of logos. These can cover anything from individuals looking for a place to shop to companies that need to make sense of their incoming mail. Logos are made for these end users; they are designed to be easily identifiable and are intended to increase recognizability. Consumers typically need applications that build on this fact and help them leverage the recognizability of the logos to facilitate or automate different tasks. Logo-based document classification is probably one of the most interesting and commercially exploited applications to examine in this context.

A third type of actors with a vested interest in the logo and trademark world are the government organizations and professionals that deal with the legal aspects of trademark registrations. For these professionals the major challenge is conducting searches for similar previously registered trademarks in large databases and identifying potential issues that would prohibit the registration of a new trademark. A crucial aspect of this application is the ill-defined nature of "similarity," which, legally speaking, is not restrictive to visual similarity but it also covers the semantics and typically extends to aspects like phonetic similarity of words.

This chapter presents the current state of the art methodologies and best practices using three key applications related to the above domains to provide the necessary context. These applications are trademark retrieval systems, logo recognition in

Table 18.2 Comparison of the different applications discussed in this chapter and summary of their respective challenges

Application	Input	Scale	Key challenges
Trademark retrieval systems	Logo-only images, high quality	Large number of models, typically in the 10–100 thousands	• Scalability • Difficulty to define the concept of visual similarity (as opposed to matching) • Lack of public datasets
Logo recognition in document images	Scanned paper documents, typically black and white	Small number of models, typically less than 50	• Application-driven assumptions • Scalability • Handling of noise and document artifacts
Logo detection and removal in images and videos	Unconstrained images or video sequences, typically colored	Average number of models, typically in the few hundreds	• Generalization • Features applicability • Scalability

document images, and logo detection in real scenes (images or video). The selection of these three applications has been made so that all different aspects of logo and trademark recognition are covered as shown in Table 18.2.

Trademark Retrieval Systems

This section outlines the evolution of trademark retrieval systems over the past 20 years and details the state of the art in the topic. First, a problem definition is attempted and the key requirements for a trademark retrieval system are identified. Subsequently, frequently used trademark descriptors are detailed, before the focus is shifted on the aspect of retrieval itself and especially on the different ways to combine results obtained over diverse modalities and on tackling the issue of subjectivity through relevance feedback techniques. A summary of a limited number of selected representative systems is finally presented through a comparison table as well as a summary of the open challenges in trademark retrieval.

Problem Definition

In most countries, trademarks must be formally registered with the national trademark office to gain legal protection. The objective of the registration process is to ensure that the trademark is sufficiently different from previously registered ones

to avoid confusion within the particular market segment(s) it addresses. The process of registering a new trademark is not trivial and can take more than a year before an application is accepted.[1]

Although different countries have slightly different rules and timelines, a key step of the process is invariably the examination of the application by an assigned attorney, an intensive process involving various searches in the registry to identify "similar" previously registered trademarks. Similarity in this context is defined on the basis of common sense. For example, in Indian law, Section 1(h) of the Trade Mark Act 1999 defines a "*Deceptively Similar mark*" as "*... so nearly resembling another mark as to be likely to or cause confusion.*" In the UK law, Section 5(2) of the Trade Marks act 1994 states "*A trademark shall not be registered if because it is similar to an earlier trademark [...], there exists a likelihood of confusion on the part of the public, which includes the likelihood of association with the earlier trademark.*"

This likelihood of confusion must be appreciated globally taking into account all relevant factors and every possible interpretation of a trademark image, "*including a visual, aural and conceptual assessment.*"[2] The legal definition of trademark similarity leaves a lot of space for subjective interpretation, frequently leading to ambiguous decisions.

In order to facilitate the searches, the trademark offices maintain details of all registered trademarks. Text-only trademarks are generally easy to compare and retrieve based on edit distance measures [47]. An extra complication arises from textual (word-only or device-and-word) trademarks which exhibit aural (phonetic) similarity for which special analysis is required [30].

Nevertheless, the major challenge, and main focus of this section, comes from graphical (also called figurative, device-only, or form-only) or combined textual and graphical (also called device-and-word or form-and-word) trademarks, which represent the majority of received trademark applications. For such trademarks, visual similarity is of prime importance. Figurative trademark searches are principally dealt with using predefined classification codes, the most widely adopted system being the Vienna classification [87] developed by the World Intellectual Property Organization. Such schemes address conceptual similarity, but no visual similarity.

An application very close to trademark retrieval is the so-called trademark watch service, where firms with registered trademarks are actively monitoring all new registrations of trademarks to make sure that they oppose to the admission of any trademarks similar to theirs. The key difference is that the large-scale retrieval process is substituted with a relatively small-scale comparison process, where new trademarks are assessed as similar or non-similar to a limited set of trademarks "under watch." The main focus of this section is on trademark retrieval, but virtually all methodologies discussed here can be applied to trademark watch.

The application of logo retrieval for the trademark office is defined as a retrieval problem, where the objective is to find all similar, previously registered figurative trademarks given a query figurative trademark. The query is a single trademark image, which is usually a high-quality scan, or even a born-digital image that contains only the trademark in question; a few systems also tackle hand-sketched

query images. The samples in the collection of previously registered logos can also be assumed to be high-quality scans or born-digital images. Metadata are available both for the query and the collection samples and correspond to a list of Vienna codes. "Similarity" in the trademark retrieval problem should be considered at as many levels as possible covering all possible trademark interpretations, usually considering shape similarity, color similarity, and semantic similarity (typically based on the Vienna codes). Trademark retrieval systems aim to facilitate the human examiner by limiting the number of trademarks to check; hence, a high precision is important to ensure their usefulness. From the legal point of view though, achieving a high recall rate is paramount as all potentially conflicting trademarks should be retrieved and checked against.

Trademark Descriptors

Searches for similar logos include a visual, aural, and conceptual assessment. The technical focus of this section is on visual similarity, although similarity over different domains is indeed discussed to the extent that it is important in the system design, as ultimately systems able to fuse information obtained over different domains are needed.

Searches for similar logos are usually performed over some notion of visual similarity, typically combining features over different modalities, primarily shape and color. Frequently, a core shape description is complemented with features related to other trademark properties, including color features, geometric features, topological information, texture features, and metadata such as semantic category codes. There is no single best-performing combination of features to use, as the importance of each visual modality to assess similarity depends on the particular design and intended use of the query trademark in question. In this section, the key visual descriptors used for assessing visual similarity are discussed as well as the use of semantic category codes as a complementary source of information.

Perceptual Organization

Depending on the retrieval solution, trademarks are described at various levels of detail. Systems that extract features from individual connected components are common [39], while methodologies that encode the topological relationship between components [2] or systems that include certain preprocessing (e.g., filling [43] or dilation [1]) in order to eliminate fine details of the trademark have also been proposed. These last methodologies stem from the observation that "*the end-user's recollection of a figurative trademark is of a general and hazy nature*" [43].

At the semantic level what is ultimately sought is an interpretation of the trademark image. This is naturally a difficult proposition for automated processes, more so as there are cases where more than one valid interpretation can be arrived at, while interpretations are usually context related. Nevertheless, it is a fact that human observers interpret the trademark image as a whole. After observing trademark examiners in action, Eakins [26] suggests that the examiners must identify and

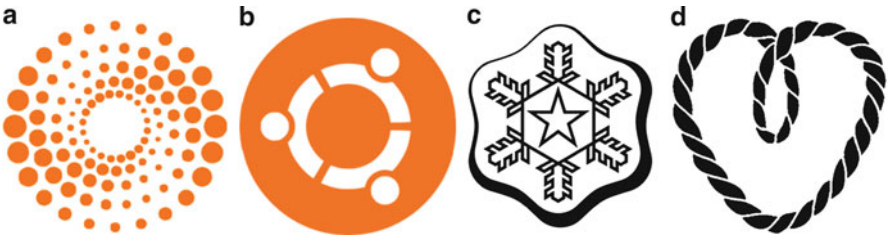

Fig. 18.1 Examples of bi-level figurative geometric trademarks. One distinctive feature of (**a**) and (**b**) is the central circular pattern, no matter whether it is made up by *dots* or *arc* segments. In (**c**) the most distinctive feature is the overall *star* arrangement of the trademark, while in (**d**) the single shape that stands out is the shape of a *heart*

remember the most distinctive feature of the query image, which might be a single large shape or a group of objects making up a recognizable pattern (see Fig. 18.1).

In order to arrive to these higher-level interpretations, an automatic system has to be able to construe certain properties of basic structural elements in the image such as symmetries, continuity, proximity, and colinearity. This is a necessity for the considerable portion of trademarks that comprise abstract geometric designs (see Fig. 18.1) but less so for other pictorial trademarks. This functionality corresponds to an early stage process of human vision, called "visual grouping" or "perceptual organization." An important function of perceptual organization is to distinguish as accurately as possible between accidental structures in the image and significant structures unlikely to have arisen by accident.

Based on the seminal work of Gestalt theorists [85, 86], a number of attempts have been made to provide computational models for the role of perceptual organization in the functioning of vision. These include early work by Lowe [51], Sarkar and Boyer [70], Amir and Lindenbaum [6], Saund [71], as well as more recent attempts notably by Desolneux [21]. The interested reader is referred to the above referenced work for technical details on the underlying computational models.

Various perceptual organization models have been used in the context of trademark shape analysis and are all viable options for extracting higher-level groups of basic elements for subsequent trademark representation. There is little comparative analysis of the different models in the trademark retrieval problem and no established best practice. See, for example, Alwis and Austin [5] who compared the bin-voting method suggested by Sarkar and Boyer [70] to the pair-wise feature extraction process of Lowe [51] with mixed results. The selection of the best perceptual organization computational model depends on a number of system design decisions.

Perceptual organization can arise over different primitives from pixels to lines and components and over a varied set of properties such as shape similarity, color similarity, parallelism, orientation, continuity, and, importantly, proximity between the primitives considered. There is great variability to the choice of

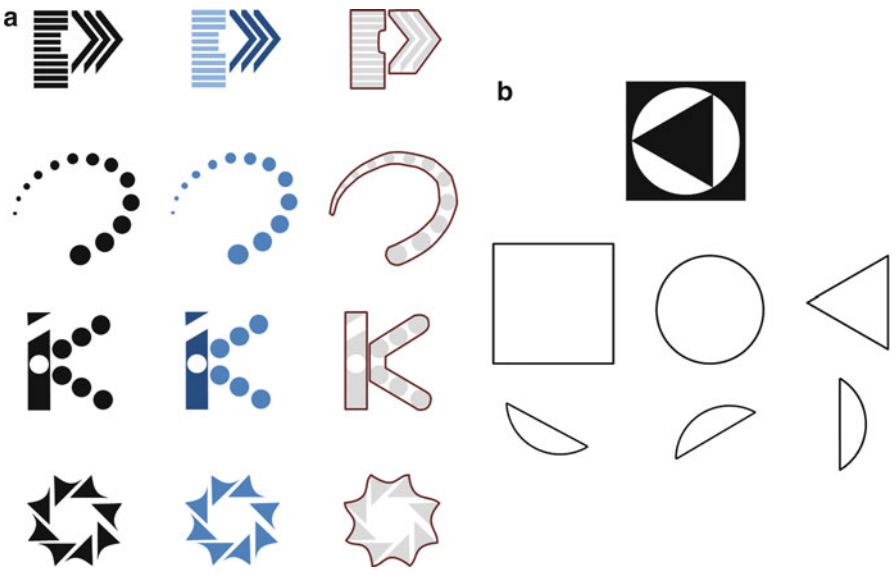

Fig. 18.2 (**a**) Perceptual organization modeled on connected components and resulting envelopes (Adapted from [1]). (**b**) Perceptual organization modeled on lines and curves and resulting closed curves (Adapted from [5])

primitives and associated properties to use. Alajlan [1] opts to perform connected component grouping over four properties: proximity, area, shape, and orientation, using the evidence accumulation framework to decide on the final groupings. Alwis and Austin [5] base the perceptual organization process on straight-line and arc segments and calculate pair-wise perceptual relationships between them such as end-point proximity, orientation similarity, curvature similarity, parallelism, colinearity, and co-curvilinearity. In the ARTISAN system [25–27], trademarks are expressed as a set of closed region boundaries which are subsequently examined to discover families of similar regions based on conditions that include close physical proximity of the boundaries, significant lengths of the boundaries being collinear or parallel, symmetry or shape similarity of the corresponding regions, and boundaries enclosing areas of similar color or texture.

When working at the level of closed boundaries of connected components, the final groups are usually approximated by an envelope shape enfolding the grouped regions (see Fig. 18.2a). When working at the level of lines or curves, the resulting groups are usually closed boundaries that represent alternative interpretations for the trademark (see Fig. 18.2b). In both cases the resulting groups are the basis for subsequently calculating shape descriptors. Consequently, indexing and retrieval is based on a combination of individual components, detected groups, and whole-image descriptors.

Although perceptual organization is accepted as an important stage in trademark image understanding, the performance of current computational models varies a

lot depending on the trademark image. Ren et al. [63] performed a comparison between human-obtained and ranked segmentations of trademark images into their constituent parts and the segmentations produced by the ARTISAN system. They report an average performance of about 58 % of the automatically produced segmentations matching within one grouping mistake one of the top human-produced segmentations. About 28 % of the automatic segmentations were in a perfect agreement with the humans' first ranked choice. A qualitative analysis of the failure types shows that the most important source of confusion is the existence of textured areas, while over-grouping was also the cause of discrepancies in numerous cases.

Apart from perceptual organization computational model shortcomings, it is worth noting that perceptual organization approaches implicitly assume that trademark images comprise a number of easily separable parts. This is indeed the case for a sizeable portion of figurative trademarks registered with the exception being photographic trademarks which cannot be treated in this way. Overall, perceptual organization is considered an open challenge, while the domain of trademark retrieval presents a very interesting test bed for novel methodologies.

Visual Similarity: Shape and Topology

The main modality over which visual similarity is assessed is the trademarks' shape. During the past two decades, virtually every type of shape descriptor has been used for trademark retrieval at various detail levels from individual trademark parts, to "envelope" shapes grouping a number of separable parts, to global shape descriptors over the whole trademark image.

Simple geometric features such as aspect ratio, circularity, compactness, central moments, Euler number, and eccentricity as well as Rosin's features (triangularity, rectangularity, and ellipticity) have been used extensively [2, 13, 27, 28]. These are a common choice when description of individual parts is sought, as they are simple and fast to calculate. More complex contour- and region-based shape descriptors have also been broadly used, including among others Fourier descriptors [27, 84], Zernike moments [45, 84], Angular Radial Transform [27], Hu moment invariants [27], Curvature Scale Space [27, 84], shape context [11, 66, 67], grid-based pixel density [25], and edge direction histograms [43]. The reader should refer to ▸Chap. 16 (An Overview of Symbol Recognition) for a detailed account of cited shape descriptors.

Eakins et al. [27] performed a comparative retrieval study between whole-image versus part-based trademark image matching and found strong support for part-based matching being significantly more effective than whole-image matching. A number of interesting observations can be made on this topic. First, the descriptors used for part-based matching are usually simple geometric features, which are much faster to calculate than whole-image descriptors. This counteracts to some extent the increase in computational complexity occurred due to the need to index and search a much larger number of parts compared to whole images. A downside of part-based approaches is that they implicitly target abstract geometric trademarks comprising

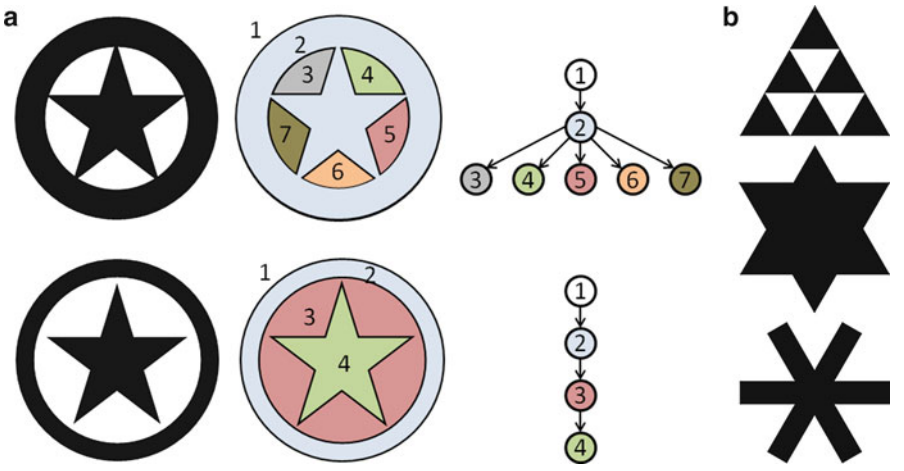

Fig. 18.3 (**a**) Two very similar trademarks with quite distinct internal topologies. (**b**) Example of different trademarks with virtually the same edge direction histograms, peaking at 0°, 60°, and 120° (Adapted from [45])

a number of easily separable parts, rendering them difficult to apply in the cases of photographic trademarks.

Regarding the efficiency of different shape descriptors, direct comparisons between published methods are difficult to perform as authors rarely use the same datasets or the same performance evaluation metrics. Nevertheless, comparisons performed independently by various authors on their respective datasets converge to similar conclusions [27, 84], also confirmed independently of the trademark retrieval applications [90]. Based on whole-image matching performance, Zernike moments seem to perform consistently better for trademark retrieval than Curvature Scale Space (CSS), Fourier, or Hu moment invariants which yield similar performances. It is also a common finding that an early fusion of different shape descriptors results to higher discriminating power and an increase in performance (see, e.g., Fig. 18.3 for the shortcomings of using single descriptors).

A number of authors have proposed ways to bridge the local (usually part based) and global description levels. Alajlan et al. [2] use curvature trees to represent the topology of trademark parts (foreground) and holes (background) and report better accuracy than employing single global-shape descriptors. Nevertheless, they observe that the topology of trademark images can be quite unstable and change due to noise, resolution changes, occlusions, etc., while there are cases where a topological description might not be a good basis for measuring similarity (see, e.g., Fig. 18.3a). Avoiding the requirement for a stable segmentation, Wei et al. [84] propose to describe the internal structure of a trademark based on statistical features calculated locally over the edge map of the image. These are combined with global shape features in a late fusion scheme. The results indicate that such an approach

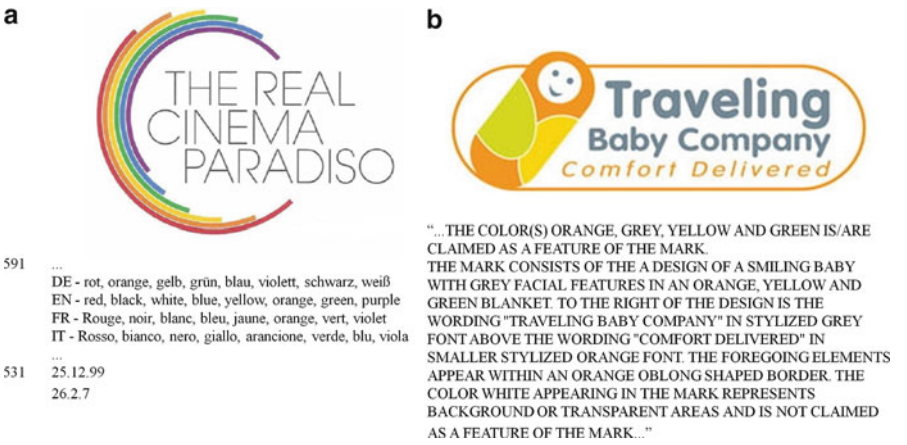

Fig. 18.4 (**a**) A trademark description as it appears in the European OHIM "Community Trademarks Bulletin," code 591 of the description corresponds to the list of *colors* in the image, while code 531 corresponds to the Vienna classification. (**b**) A trademark description in the USA "Official Gazette for Trademarks," the *color* content and its use in the image is explicitly described in the text

can yield significantly better retrieval rates than a number of global shape descriptor combinations compared against.

Techniques for shape and object recognition based on local descriptors are gaining popularity. An example of such a descriptor used for trademark matching is the shape context, proposed by Belongie et al. [11]. In [11] matching between trademarks is performed through bipartite graph matching and the Hungarian method, yielding very good results but being difficult to scale up for large databases of trademarks. Rusiñol et al. [66] made use of the shapeme histogram descriptor [56], combined with locality-sensitive hashing to achieve approximate k-NN-based retrieval in sublinear time, a methodology later extended to incorporate color information [67].

Shape is a key modality over which visual similarity is assessed. Although there is no best shape descriptor for the trademark retrieval application, there is consensus that multiple-level (part-based combined with global) descriptors work better than any single ones, while combining various shape descriptors is generally necessary, especially if diverse types of figurative trademarks are targeted (e.g., geometric and photographic).

Visual Similarity: Color Content

In the case of figurative trademarks, registering them in color can have certain legal implications. In general, logos are submitted in black and white in which case the registered trademark covers the rendering of the logo in any color combination. But if the applicant claims color as a feature of the mark, then the particular color scheme is explicitly included in the registered trademark description[3] (see Fig. 18.4).

For all trademarks that are registered in color, the specific color scheme used is intended to be a recognizable feature and strongly associated to the brand, hence important to take into account during the search (think, e.g., of the Coca-Cola red/white logo, the BMW blue/white logo, or the colored Burberry pattern). According to the OHIM manual of trademark practice[4] "*the use of the same color or color pattern may increase a visual similarity of the figurative or word elements themselves. The exact effect of colors or color patterns has to be assessed individually in each case, since this depends very much on the impact of the color on the overall impression of the signs involved.*" Therefore, as far as trademark registration searches are concerned, both color similarity and shape similarity should be taken into account as "confusingly similar" trademarks can arise in either domain.

Contrary to logo detection in videos and real scenes, relatively few trademark retrieval systems incorporate color in the trademark descriptors. Within these systems, color is variably used either as local feature [67], complementary to part-based shape descriptors [39], or as a global feature at the image level [68]. Mere coincidence of the color if the figurative or word elements are not otherwise similar is not legally considered enough to lead to a relevant similarity. It can therefore be argued that color information should be always assessed in the context of its spatial arrangement and global color features are less relevant in the trademark retrieval application.

Employing color as a feature aims to improve the retrieval of visually similar (as opposed to semantically close) trademarks; hence, it is a reasonable assumption that perceptually relevant color systems and metrics should be used to assess color differences (see, e.g., [68, 88]). This is not yet common practice, with most authors opting instead to more computationally efficient color representations.

It is also worth noting that figurative trademarks comprise highly stylized graphics, made up of few, identifiable colors. Therefore, methodologies such as color naming or other fixed color categorization schemes offer a valid alternative to color information encoding given the nature of trademark images. In this line, Rusiñol et al. [67] combine a statistical color naming model and shape context features in a late fusion scheme to perform trademark retrieval. Although color names information is included in the trademark description (see Fig. 18.4), no existing methods make use of the registered color names, bridging the metadata and image domain to enhance image-based trademark retrieval.

Overall, color information has not been extensively incorporated in trademark retrieval systems up to date, although there is a consensus that incorporating color information boosts retrieval results. Given the growing number of trademarks registered in color, this is one aspect of trademark retrieval that is expected to receive increasing attention in the years to come.

Conceptual Similarity: Semantic Categories

Another key modality in which a likelihood of confusion might arise is conceptual similarity. According to current trademark practice "*Signs are conceptually identical or similar when the two signs are perceived as having the same or a similar semantic content.*"

The main instrument available to the registration offices to detect such conceptual similarity is the Vienna classification system [87], which offers a comprehensive list of tags under different categories, adopted by trademark registrants around the world. An excerpt of the categories tree is shown below.

Category 1 CELESTIAL BODIES, NATURAL PHENOMENA, GEOGRAPHICAL MAPS
Category 2 HUMAN BEINGS
Category 3 ANIMALS
Category 4 SUPERNATURAL, FABULOUS, FANTASTIC OR UNIDENTIFIABLE BEINGS
Category 5 PLANTS
5.1 TREES, BUSHES
5.1.1 Trees or bushes of triangular shape, conical shape (pointed at top), or "candle-flame" shape (firs, cypresses, etc.)
5.1.2 Trees or bushes of oblong shape (poplars)
5.1.3 Trees or bushes of some other shape
5.1.4 Trees or bushes without leaves
...
Category 6 LANDSCAPES

Manual classification of images is time consuming and potentially error prone [26]. There is ongoing research in methodologies for the efficient use of such hierarchical classification schemes to improve retrieval. See, for example, [17] for a trademark retrieval system based solely on the semantic description of an image, where the analytical hierarchy process is used to manually assess the pair-wise relative importance of different categories within a trademark image at the time of categorization. Although such systems technically fall outside the context of this chapter, they should nevertheless be considered as baseline for evaluating CBIR approaches.

Cautiousness is advised when relying on classification codes for retrieval. A drawback identified by various authors is that classification codes are not helpful for abstract images (a sizeable portion of registered trademarks comprise geometric designs that have little or no representational meaning), while they are inadequate to describe such attributes as color, shape, texture, and layout [26, 43]. Figure 18.5 shows a number of trademarks categorized under the same code (05.01 – "Trees, Bushes"). It can be easily appreciated that conceptually similar trademarks might show no visual similarity whatsoever. Conceptual similarity might come into play for two signs as a whole or in comparing elements of composite signs. Moreover, conceptual similarity between the word and the figurative aspects of different trademarks (e.g., the word mark "tiger" against a figurative mark depicting a tiger) is also legally accepted as a basis to establish a likelihood of confusion condition and is therefore searched for by the officers. The main conclusion to draw is that semantic category tags and visual descriptors are highly complementary.

A limited number of trademark retrieval systems combine visual similarity search with semantic information to improve or refine results. In certain cases a

Fig. 18.5 Trademarks categorized under code 05.01–"Trees, Bushes" of the Vienna classification (Trademarks obtained through the eSearchPlus online service of OHIM (May 2012), accessible online at: http://esearch.oami.europa.eu/copla/advanced#/trademarks)

prior keyword-based search is assumed to have taken place to reduce the size of the search space before content-based image retrieval methodologies are applied [43]. Combining shape, color, and semantic information (afforded through Vienna codes) is attempted in [68], through the use of predefined weights to specify the relative importance of each modality. In [88] a fuzzy system is used to calculate the distances between sets of terms, which are subsequently combined with image-based descriptors.

The existence of an international classification standard should be taken into account both as a possible means for training object (concept) detectors and in its capacity as a complementary retrieval domain enabling multimodal retrieval systems.

Information Fusion, Indexing, and Retrieval Strategies

Contemplating the functioning of a complete trademark retrieval system, it is important to remember that three domains are legally defined where a likelihood of confusion might arise: visual, aural, and conceptual. Moreover, within the visual domain, which is the main focus of this chapter, numerous modalities can be taken into account, shape and color being the most prominent ones discussed here. An extra complication when comparing pairs of trademarks is that different aspects of their description, not known a priori, bear more importance than others, making it difficult to properly weight the contribution of each description modality.

Developing a working trademark retrieval system implies selecting the right strategies to efficiently fuse information from different domains. The best strategy to use is inherent to the structure and focus of each system. An overview of the main schemes along with a discussion of advantages and disadvantages is given here.

Early Fusion and Unimodal Retrieval

Approaches that rely on early fusion first extract unimodal features which are then combined into a single representation based on which classification and retrieval is

performed. The main advantage of such schemes is their computational efficiency at the time of retrieval, especially for relevance feedback techniques as a single feature vector representation is involved. Nevertheless, the meaningful normalization of features extracted over different domains and modalities (e.g., shape, color, texture, semantic codes, aural representations) is generally difficult to achieve.

Early fusion schemes frequently boil down to unimodal retrieval systems, most commonly shape similarity-based systems that combine a variety of distinct shape features.

Late Fusion and Multimodal Retrieval

Late fusion is a relatively easier way to deal with information originating over different domains. Late fusion focuses on the individual strength of separate modalities, and in the trademark retrieval application, it is usually employed in two distinct manners. Before retrieval, unimodal similarity (distance) scores are fused into a single multimodal similarity (distance) measure which is then used to rank results. Otherwise, retrieval can be performed separately using different descriptors, subsequently combining the obtained ranked lists of results.

Usually, similarity scores obtained over different modalities are fused through a weighted average scheme. Other choices include different operators (e.g., sum, multiplication, max [39]) and fuzzy systems. An early example of this practice is presented in [88]. The system enables comparisons between word marks, between device marks, and between marks of different types through a sequence of heuristics based on weighted distance calculations. For word marks, edit distance, phonetics, and interpretation scores are calculated independently and subsequently combined based on user-defined weights. Similarly for device marks, meaning (based on Vienna classification codes), structure, and shape similarities are combined. Finally, to enable comparison between marks of different types type-mismatch penalties are introduced.

A typical need of trademark retrieval applications is the combination of retrieval results (ranked lists of similar trademarks) obtained over different modalities. Using ranks for this fusion has advantages over combining incommensurates which differ in range, mean, and variance since they come from different representations and mechanisms [10]. Reciprocal rank fusion, Borda count, logistic regression, and the highest rank methods are typical choices for combining ranked lists obtained from multiple retrievals [2, 5, 39].

It is important to note that when working with semantic categories, there is no implicit ranking between trademarks that share the same number of classification codes. Instead, where a retrieval system would produce ranked lists for modalities such as shape and color, it produces separate sets of equally ranked trademarks (sharing the same number of categories with the query trademark) in the case of semantic information. For such cases, it is important to use methodologies that take care of the ties in the semantic results, such as the Condorcet method (see, e.g., [68]). A downside of this kind of methods is that they are usually computationally expensive as they require a lot of pair-wise comparisons; hence, approximate solutions or prior filtering of the results should be sought.

Multistage Retrieval

A different variety of systems employ a multi-scale retrieval strategy where a cascade of filters are applied, aiming to reduce the search space before a final ranking is calculated among the filtered trademarks. This approach is common with perceptual grouping-based methods and methods that combine multiple-level (part-based and whole-image) descriptors. It also provides an alternative way to define the relative importance of different modalities, reflected on the order of the applied filters.

Systems based on multiple-level descriptors usually employ a two-stage retrieval scheme, where global features act as a kind of hash variable to retrieve a subset of trademarks which are then compared on a part-by-part basis. As an example, in the ARTISAN system [25], a hierarchical record is created for each image with shape information stored at each level (whole image, family, boundary). Retrieval is accomplished by first comparing whole-image shape features, deemed more generic, and then traversing ordered lists of image components comparing each query component with the closest-matching stored one.

On a similar line, Hung et al. [42] calculate two shape features per trademark image, a rough contour-based descriptor and a more detailed region-based one. The database of trademark images is categorized automatically into 11 classes based on the contour feature. During retrieval, the contour descriptor of the query image is used to dynamically select the possible classes that the image belongs to, forming candidate sets with different priorities, then the region-based descriptor (Angular Radial Transform) is employed to decide on the final ranking. Jain and Vailaya [43] use a similar process, performing a rough retrieval based on edge direction histograms and invariant moments, followed by a refinement stage based on deformable template matching.

An example spanning different modalities is the methodology of Hsieh and Fan [39]. They follow a two-step similarity assessment, where first shape and topology similarity probabilities are combined, and then the result is combined with the color similarity probability. Other typical applications of multistage retrieval techniques are when semantic categorization is available and is used as a first filtering.

Relevance Feedback

The importance of different modalities in the visual similarity search is to a great degree a function of the particular trademark under question. As this is not easy to define prior to the search, certain trademark retrieval systems allow the search officer different ways to decide the relative importance of the different modalities. This is achieved either by adjusting the relative weights of different features manually or dynamically by allowing for a relevance feedback process.

Relevance feedback is a user-controlled iterative process for query reformulation. The key concept is to iteratively improve the retrieval result by emphasizing previously retrieved relevant items and de-emphasizing irrelevant ones. The two main variants work either by reformulating the query in every step (e.g., the Rocchio algorithm [64]) or by revising the similarity metric (e.g., Giacinto and Roli [34]).

Fig. 18.6 In the *top* query, the user opts to select relevant images that are visually similar to the original query. In the *bottom* query, the user opts for trademarks that are conceptually similar (they contain some a representation of a *tree*). The system adapts accordingly (Reproduced from [68])

Numerous examples of relevance feedback-based systems for trademark retrieval have been reported in the literature, including Chou and Cheng [17] who propose a semantic-only system using Rocchio algorithm and Ciocca and Schettini [18] who propose an algorithm that updates both the weights used to combine the different shape features as well as reformulate the query accordingly in every step.

Applying relevance feedback over three modalities independently, Rusiñol et al. [68] propose a framework that combines shape, color, and semantic information (Vienna codes). The system considers the three modalities independently, producing ranked lists for shape and color and separate sets of trademarks sharing the same categories in the case of the semantic information. These are combined using the Condorcet method in order to take care of the ties in the semantic results. A relevance feedback mechanism then allows the user to shape the query in real time. In reality, the three queries (independent modalities) are independently updated and the ranked results are fused as above in each step. Two typical retrieval scenarios are shown in Fig. 18.6.

Relevance feedback is an important addition to trademark retrieval systems, accepting the limitations of automatic descriptors in covering the whole space of possibilities and the importance of expert's feedback to the search process. Relevance feedback allows the expert user to direct the search towards the desired outcome focusing on visual or conceptual similarity, as required by the trademark under question, on the basis of his iterative feedback.

Systems Overview

Table 18.3 attempts to provide a summary of the key aspects of selected methods spanning the past 15 years of research in trademark retrieval systems. This is not supposed to convey a complete list of trademark retrieval systems but instead

Table 18.3 Comparative summary of different trademark retrieval systems

Name	Year	Description level	Shape descriptor	Color	Semantic categories	Perceptual organization	Dataset	Relevance feedback
Wu et al. [88]	1996	Part-based	Fourier coefficients, moment invariants, projection histograms	N/A	Vienna codes, processed into a Fuzzy thesaurus	User-assisted grouping	3,000 color trademarks, scanned in-house	N/A
Kim and Kim [45]	1998	Image level	Zernike moments	N/A	N/A	N/A	3,000 bi-level trademarks, scanned in-house	N/A
Jain and Vailaya [43]	1998	Image level	Histogram of edge directions, moment invariants, deformable template matching	N/A	N/A	N/A	1,100 bi-level images, scanned in-house	N/A
Alwis and Austin [5]	1998	Multilevel	Graph-based representation, parts described by primitive features	N/A	N/A	Pair-wise feature extraction (Lowe and Sarkar/Boyer)	1,000 bi-level trademarks, subset of [26]	N/A
Hsieh and Fan [39]	2001	Part-based	Fourier coefficients, topological relationships between parts	RGB or HSI	N/A	N/A	155 color trademarks, scanned in-house	N/A

Eakins et al. [25–27]	2003	Multilevel	ART, moment invariants, Fourier, CSS, Rosin descriptors	N/A	N/A	Boundary families based on proximity, symmetry, and colinearity conditions	10,745 bi-level trademarks from UK trade mark registry	N/A
Alajlan et al. [2]	2006	Multilevel	Curvature trees, Triangle Area Representation (TAR)	N/A	N/A	N/A	1,500 bi-level logo images (400 from MPEG-7 CE-2 database, 1,100 from [43])	N/A
Alajlan [1]	2007	Multilevel	Eccentricity, solidity, orientation for the parts. TAR for final envelopes	N/A	N/A	Hierarchical clustering and evidence accumulation	110 bi-level trademarks	N/A
Wei et al. [84]	2009	Image level	Zernike moments, edge curvature, and distance to centroid	N/A	N/A	N/A	1,003 bi-level trademarks, 14 classes	N/A
Rusiñol et al. [67]	2010	Image level	Shape context	Color names histogram	N/A	N/A	323 color trademarks, in 24 classes	N/A
Rusiñol et al. [68]	2011	Image level	Shapeme histograms	CIE LCh color histogram	Vienna codes	N/A	∼30,000 color trademarks from the Spanish IP Office, 1,350 Vienna categories	Rocchio

to offer a representative selection of methods that cover in various combinations all the topics discussed in this section. It is extremely difficult to provide any direct comparisons between the different systems exposed as both the datasets and evaluation metrics used vary widely.

Open Challenges in Trademark Retrieval

For a trademark retrieval system to be practical and commercially successful, it has to be able to work in real time, achieve a recall rate close to 100 % (no potentially similar trademarks missed) with a high precision (minimum number of irrelevant trademarks shown to the officer), and allow for variable definitions of similarity that span different domains. A number of open challenges stand on the way to achieve this, a summary of which is attempted next (see also [73] for an interesting discussion).

Scalability and Variability Issues

The number of trademarks registered worldwide is estimated to exceed 20 million. The majority of registered trademarks are actually word-only ones, substantially bringing down the number of device, and word-and-device ones, which are the main focus of this chapter. From the point of view of the collection size, this number of samples is manageable. Nevertheless, when considering that multiple retrieval processes might be necessary to tackle different modalities or when considering methodologies based on large numbers of local descriptors extracted from each image with their associated matching processes, it becomes obvious that a number of the techniques currently proposed would not scale gracefully with the size of trademark collections.

The variability of trademarks is another source of challenges. Trademarks are formally classified as word-only, word-and-device, and device-only ones (leaving out special trademarks like 3D, smell, and color per se) depending on their content. But within these categories there is substantial variability including geometric versus photographic trademarks, monochrome versus color trademarks, and stylized text versus plain word-only trademarks. It is commonplace that individual methods put forward only address a particular subtype of logos.

Availability of Resources (Datasets and Ground Truth)

It might come as a surprise that although an intricate international structure exists to register and control trademarks, it is relatively difficult to encounter publicly available collections of trademarks. Although trademarks are available to browse online one by one, they are not available to download and use as a collection. As a result, the use of private trademark datasets is widespread.

A consequence of the lack of any standardized dataset is the lack (and difficulty to obtain) of associated ground truth data to train and evaluate trademark retrieval systems on. A direct way to evaluate such methods would be to obtain and compare against human expert responses given the same test queries and dataset. A few

authors attempted this with limited data [13, 43], but the sheer amount of time required to obtain such responses makes this unpractical. Jain and Vailaya [43] report that it took each subject between 1 and 2 h to select the top ten similar results for 5 query images, within a dataset of 1,100 geometric-only trademarks. In the absence of any (easily obtainable) ground truth, performance evaluation is often based on a posteriori assessment of the returned queries, which gives a subjective qualitative measure but no quantitative evaluation.

Visual Similarity and the Importance of User Feedback

Visual similarity is an ill-defined concept, as it can arise over different attributes for any given pair of trademarks. There is no way to define a priori the relative importance of the different modalities for assessing visual similarity. Two common practices exist to partially alleviate this problem. On one hand, a trademark retrieval system should take into account all possible interpretations of a trademark image. This entails performing multiple visual similarity checks over different modalities or over alternative descriptions (e.g., different groupings of trademark parts) and either combining their results or directly presenting alternative results to the user. On the other hand, relevance feedback methodologies let the user implicitly decide over what modalities search should be performed, iteratively adjusting the relative significance of different attributes.

A couple of observations can be made here. First, it seems that the increased participation of the user is necessary. This opens up new problems both in terms of user interfaces (see [80]) and on how to best learn from the user feedback, what calls for interdisciplinary research. Schietse et al. [73] suggest various ways to achieve the improved involvement of the user, including the ability to specify how the query should be formulated (e.g., based on a complete image or on specified parts, define the search parameter weights such as shape and color), relevance feedback, and improved display paradigms.

Second, the trademark offices' databases contain a lot of metadata information that is not efficiently used by current methods. For example, the fact that a trademark is submitted in color is important as it signifies that color has an increased significance in this particular case. Future systems should examine how this information can be best used to fine-tune the searches. Use of metadata information can be made both for learning and for filtering and improving results presentation. The existence of millions of logos classified under the Vienna classification is a valuable resource that could drive graphics recognition and training of complex object detectors.

The Difference Between Similarity and Matching

The aim of trademark retrieval systems is to identify visually similar trademarks as opposed to perform exact image matching. The difference is subtle and should indeed boil down to relaxing sufficiently the similarity thresholds used. Nevertheless, many of the trademark descriptors employed were never designed to be used like this; instead they were designed to be robust in an image-matching scenario and have a sharp response as similarity decreases in order to reduce false positives. Another way to say this would be that depending on the trademark descriptor, small

changes in the distance threshold do not necessarily correspond to a human notion of smoothly decreasing similarity.

Future Outlook

Overall, trademark retrieval, as well as trademark watch, is an open area for research with clear commercial value. Most existing research on CBIR is focused on natural images. Techniques for trademarks, but also for other artificially produced images such as icons, logos, coats of arms, and clip-art images, have received less attention, even though there is evidence that these images require a different set of techniques for effective retrieval [73].

The topic of trademark retrieval offers a lot of opportunities for future research. Apart from document analysis-related research, looking into better and more perceptually relevant shape, texture, topology, and color representations for this kind of images, there is also a wide scope for multidisciplinary research. A non-exhaustive list of research areas with potential impact includes the enhanced involvement of the user, scalable indexing and retrieval methodologies, the efficient use of information from nonimage domains, and research on better perceptual organization models.

Logo Recognition in Document Images

This section discusses methodologies for the detection and recognition of logos in document images. It also highlights the usage of logos for document classification and retrieval purposes. First, a problem description is offered. The section continues with a discussion on the detection and localization of logos in documents, followed by the particular use of logos for document classification. Finally, a summary of the open challenges in logo detection in document images is given.

Problem Definition

Logo detection and recognition in document images is of great interest for automatic document processing, since logos are indicative of the document source [91]. As such, logo detection and recognition facilitates several aspects of the document processing pipeline as it offers a viable solution to document classification and indexing (see also ▸Chap. 7 (Page Similarity and Classification) for alternative methodologies to page classification). To illustrate the economic importance of such tasks, it suffices to at digital mailroom implementations. It is estimated that companies receive on average three million documents per year, more than half of which originate in paper. Based on recent market research in the UK, the cost to classify and archive a document is circa £1, while the cost of searching for a document ranges between £5 and £300.[5] In this section the different logo-related processes involved in such tasks are discussed, namely, logo detection, logo recognition, and logo spotting.

Logo detection refers to the process of asserting whether a document contains areas that could possibly be logos and localizing such areas. Generally, no specific a priori defined logo models are assumed, instead global logo characteristics are employed. Logo recognition then refers to the subsequent process of classifying a logo candidate area to one of a list of predefined logo models.

Logo spotting on the other hand refers to the process of detecting whether a predefined query logo appears in the target document. Logo spotting typically involves an off-line processing stage where incoming documents are analyzed and indexed based on local features obtained over identified keypoints or key regions. Following this, given a query logo, spotting becomes a retrieval exercise in the indexed keypoints (or key region) domain. Given the recognizability-by-design of logos, searching for documents cast as a logo-spotting exercise that proves to be a highly effective way to retrieve relevant documents.

Real-world applications feature continuous document flows that need to be processed in real time. The time allocated to document categorization is limited; hence, a key aspect of all above tasks is their ability to perform in a close to real-time manner. This issue is typically manifested as a scalability requirement, where proposed methodologies should be easily scalable both in terms of the number of documents processed per unit time and in terms of the number of logo models considered.

Detection

Several methods in the literature related to logos in document images deal with the logo recognition problem assuming that the detection of logos has been performed by a separate module [15, 22, 35, 57]. On the other hand, there are methods that attempt to detect and localize instances (if any) of a logo in a document page. One way to distinguish these methods is according to the analysis level used for the description of the document's content. There are two levels of description. The first one is based directly on the pixel color information without any preprocessing, while the other involves some kind of preprocessing or layout analysis in order to extract higher-level structures from connected components or regions.

Pixel-Based Detection

Approaches in this category do not involve any preprocessing step, and the focus is on finding pixel-based local structures that are likely to be logo candidates. For example, in [61] a training-free unconstrained logo detection method is described based on the principle that the spatial density of the foreground pixels within a given windowed image that contains a logo is greater than those of non-logo regions. Each pixel is considered as a potential cluster center of a windowed area, and its spatial density is computed in order to decide if it is a logo or not.

In a similar way, Wang and Chen [81] proposed a method based on the assumption that almost all documents with logos keep a relatively larger distance for the logo from other parts in the documents. In their algorithm, each foreground

pixel is taken as a seed feature rectangle which is growing in order to embrace the logo-candidate area. A tree classifier is then employed to quickly discard the candidates whose likelihood to be logos is very small.

The advantage of the pixel-based approach relies on its robustness in case where the content of the document image is subject to small changes of low variability. Moreover, pixel level techniques may be efficiently implemented by fast image processing operations that are benefiting by the local distribution of pixels (windowing approaches, parallel calculations, etc.)

Region-Based Detection

Instead of using the pixel information directly, there are several methods where the binarized image is segmented either into labeled areas (e.g., text, image, graphics) or into a number of connected components which are then described by properly defined features. For instance, Zhu and Doermann [91] proposed a multi-scale boosting approach that involves Fisher classifiers in order to detect logo candidates in a document image. An initial two-class classification is performed at a coarse image scale on each connected component. An identified logo candidate region is successively classified at finer image scales by a cascade of simple classifiers, which allows false alarms to be quickly rejected and the detected logo to be more precisely localized. The context distance, the spatial density, the aspect ratio, and the area are some of the features used to describe the connected components.

In an early approach, Seiden et al. [74] considered the problem of classifying segments of a binary document image according to whether or not they are likely to contain a logo. Segmentation is performed using a top-down, hierarchical X-Y tree scheme. For each segment a set of 16 features are derived that are based on statistics about the connected components of black pixels within the segment. A subset of segments is used as training set from which classification rules are derived based on the C4.5 algorithm. The rules are then used to classify the remaining segments based on whether or not they are likely to contain a logo or not.

Overall, if layout analysis or segmentation into connected components can be efficiently applied on the document under consideration, then this approach is advantageous since the extracted higher-level structures have a semantic importance that is meaningful and can be further explored. However, region-based detection methods are intrinsically prone to any errors inheriting from the layout analysis or the connected component extraction step.

Recognition

Given a detected logo candidate found in a document image, logo recognition attempts to classify the logo as belonging to one of a finite number of logo classes or conclude that it does not belong to any known class. The efficiency of the recognition system relies on the discriminative power of the features that describe the logo candidate image as well as on the robustness of these features on distortions and noise. Virtually all the descriptors discussed previously can be applied to

recognition of logos in documents. However, in contrast to the trademark retrieval process where the goal is asserting visual similarity (trademarks spanning various classes are expected to be retrieved), here the focus is on recognition, namely, on deciding on a specific logo class given the detected candidate area.

Fusion of Structural and Topological Features

A characteristic that distinguishes a logo from other visual objects is that it consists of several parts whose structure and topological relations is important for the description of the logo. Feature-based approaches are characterized by high sensitivity to the features selected for representation and are unable to represent any specific information about the structural relationships among components. Therefore, a desirable property is the ability of a method to take into consideration the local structure information as well as to handle structural changes. Several authors tackled this problem by proposing extensions of classical Artificial Neural Networks (ANN). For example, Diligenti et al. [22] proposed to represent the input patterns by means of directed ordered acyclic graphs where the nodes may contain either symbolic or real-valued features. A properly extended version of the classical multilayer perceptron called backpropagation through structure is also involved so as to take into account both the numerical and the structural nature of the given input graph.

A modified learning algorithm for an ANN called edge-backpropagation is also used in [35], which considers a new weighting error function. The method tries to overcome the problem of spots, stripes, and ink blobs that are present in real-world documents. The weights used in the computation depend on the gradient of the image so as to give less importance to uniform color regions, like the spots. As a result, the edge-backpropagation ANN parameters are updated by taking into account only the part of the logo which is clearly readable, and the network learning capabilities are not wasted in the reproduction of the noise.

Multimodal Approaches

A number of authors highlighted that a multilevel approach is beneficial for the recognition of logos; however, it is difficult to a priori train such a system. Therefore, depending on the applications, there are logos better described in global level while others are benefiting by a partial-based description. A hybrid method that uses both approaches in an adaptive way is given by Neumann et al. [57]. In this work, three different approaches for classifying logos are examined. A local method computes statistical and perceptual shape features for each connected component. A global method uses a wavelet decomposition of the horizontal and vertical projections of the image. Finally, a hybrid approach involves adaptively defined weights that specify the relative importance of each method according to an estimate of their relative performance.

In Doermann et al. [23] a hierarchal approach is used to prune the searching space in order to match a candidate logo using different features at the different stages of the approach. A combination of text, shape, algebraic, and differential invariants is involved. The algebraic invariants handle cases in which the whole shape of the logo

is given and it is easy to describe. The differential invariants cover complex arbitrary logo shape and handle situations in which only part of the logo is recovered.

A multi-scale strategy is also adopted in [91]. The initial classification is performed on each connected component using the Fisher classifier at a coarse-scale level, regarding the logo region as a gray-scale blob. Each logo candidate region is further classified at successively finer scales by a cascade of simple classifiers.

Logo Spotting

Another way for detecting and recognizing logos in documents is based on interest point detection and extraction of features. In recent years there has been growing interest in detection and recognition models that use local image features. These features are calculated at particular interest points and are typically characterized by scale and rotation invariance as well as robustness to noise and to changes in illumination. The difference compared to the aforementioned methods is that in approaches based on local image features, the logo model is given and the methods attempt to localize instances of the model (if any) in a document page. As an example, local features combined by a set of spatial coherence rules are used in [65]. A logo model is spotted inside the document image, and in addition the category of the queried document is also determined. The document categorization and the logo detection are performed in a training-free fashion by using a bag-of-words model of visual words. These visual words are described by two types of local features, namely, the SIFT features [50] and the shape context descriptor [11]. Figure 18.7 depicts an example of the feature-matching process between the local features of the query logo and the document image.

Instead of using classical keypoints detectors, connected components can also be used in order to extract key structures. Li et al. [48] proposed such a method where a logo is defined as a group of features with restrictive geometric relationships. For every connected component, a descriptor is calculated based on geometric and statistical properties that are related to its convex hull and are invariant to image scaling and rotation.

Systems Overview

In Table 18.4 a comparative summary is provided that examines selected methods for logo recognition in document images.

Open Challenges in Logo Recognition in Documents

Assumptions and Heuristics

It is not uncommon to adopt certain assumptions in order to facilitate the detection and recognition processes and improve the overall performance. For example,

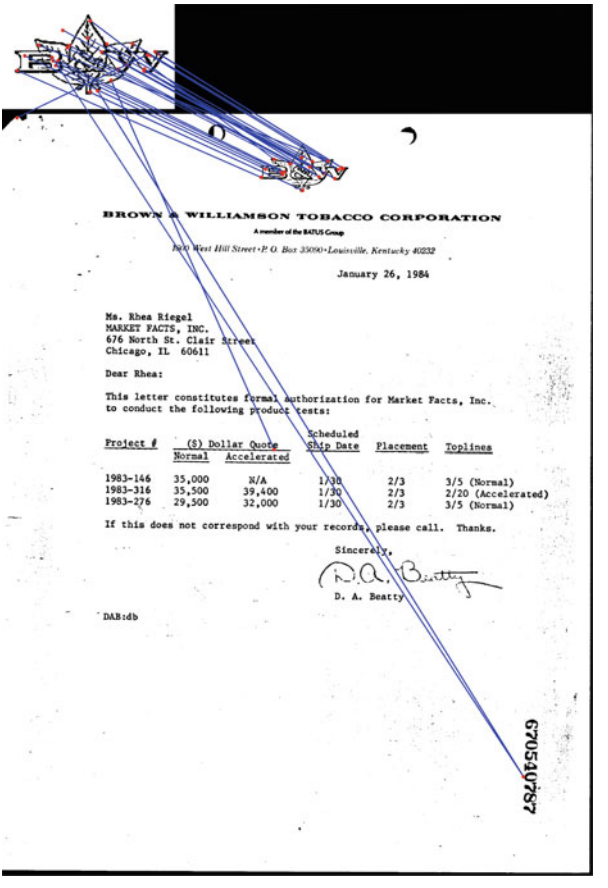

Fig. 18.7 Feature matching between a logo model (shown in the *upper left* corner) and the complete document image

prior knowledge is commonly used regarding the logo size and aspect ratio for prescreening. It is also often assumed that the logos are located at certain parts of the document, e.g., on the top one-third of the document, or in a relatively larger distance from other parts of the document. The number of logos in a document image is another issue that is usually addressed by considering only one logo per image. Alternatively, if more than one logo is detected, then only the one with its projection furthest away from the decision threshold is considered. Such implementation heuristics are commonly applied for speeding up the computational execution time. There are also methods where part of the computational cost is transferred to a preprocessing step. For example, before the matching process is initiated, the logo instances undergo several transformations in order to achieve translation, rotation, and scale invariance.

Overall, the large intra-class variations of logos and the diverse quality and degradations in captured document images increase the difficulty of the logo

Table 18.4 Comparative summary of methods for logo recognition in document images

Name	Year	Logo segmentation required	Features + descriptors	Detection + recognition	Dataset
Doermann et al. [23]	1996	No	Text + contour features, similarity invariants for unrecognized contours	Database pruning by recognized characters and shapes	University of Maryland logo database
Seiden et al. [74]	1997	No	16 features of connected components statistics	Training with logos subset + C4.5 classification	University of Maryland logo database
Diligenti et al. [22]	2001	Yes	Symbolic or real-valued features represented as directed ordered acyclic graphs	Backpropagation through structure	University of Maryland logo database
Neumann et al. [57]	2002	Yes	Local negative symbols + global wavelet decomposition	Ranking by adaptively weighted contributions of local and global methods	University of Maryland logo database
Chen et al. [15]	2003	Yes	Normalized line segments of contours	Similarity of line segments + modified Hausdorff distance	University of Maryland logo database
Gori et al. [35]	2003	Yes	Gradient of image pixel values	Modified edge-backpropagation	University of Maryland logo database
Pham [61]	2003	No	Spatial density of foreground pixels	Maximizing the mountain function	University of Maryland logo database
Zhu and Doermann [91]	2007	No	Context distance, spatial density, aspect ratio, and area	Multi-scale simple classifiers	Tobacco-800 dataset
Rusinol and Llados [65]	2009	No	SIFT, shape context descriptor	Bag-of-words approach	18 logos + 1,000 real document images
Wang and Chen [81]	2009	No	Candidate logo areas defined by boundary extension	Tree classifier	Tobacco-800 database
Li et al. [48]	2010	No	Geometric and statistical properties of connected components + line profiles	Matching to database features using the anchor line	Tobacco-800 database

detection problem. Therefore, most researchers make certain assumptions in order to have their system work, and the main challenge is to eliminate these assumptions as much as possible.

Distortion and Noise

Logo detection and recognition becomes a demanding issue due to the diverse scanning qualities, distortions, and noise that usually appear in document images [35]. In [15] Chen et al. provide a list of such distortions that includes scaling, rotation, random broken lines, random added strips, occlusions, Gaussian noise, skew, component editing (adding, removing, or skewing the logo component), as well as photo deformations such as pinch, punch, sphere, twirl, and ripple. Besides the aforementioned distortion types that refer to documents in general, there are special degradations that obstruct the document in unpredictable positions, changing the visual appearance of the pictures significantly. The most common are [22] (a) stripe noise, which is created by randomly positioned stripes of varying thickness on the logo; (b) blob noise, which is produced by means of circular isolated blobs with random center and radius; and (c) spot noise, which is obtained by superimposing many circular blobs so as to produce complex shapes that look like ink blots.

These are quite usual in real-world documents due to the common document processing chain. For example, faxes can show stripes due to transmission errors or dirt in the equipment; ink blobs or lines can be inserted accidentally or willingly to add notes or special signs to the document. The spots can change drastically the structure of a symbol (e.g., by closing a hole or connecting two or more distinct subparts), thus making the recognition with structural or syntactical methods very hard. Figure 18.8 depicts some noisy logo instances on real documents as well as their modeled counterparts.

In general, the content of documents generally includes a mixture of machine-printed text, diagrams, tables, and other elements. An efficient logo detection method must consistently output complete logos while attempting to minimize the false alarm rate. Additionally, the accurate localization needed for logo recognition poses another major challenge.

Scalability

As Tombre and Lamiroy pointed out in [77], the system's scalability is, in general, one of the main concerns in the symbol and graphics recognition domain. Indeed, in recognition schemes that rely on a learning stage and a classification strategy for the recognition of input logos, it is common to observe a decrease in the recognition rate as the number of considered model classes grows.

On the other hand, there are scalability issues regarding the number of documents processed per time unit. The ability to robustly detect logos in large volumes of documents is pivotal for logo recognition. Therefore, testing the detection and recognition methods on large document collections in order to demonstrate achievable generalization performance is an ongoing challenge [23].

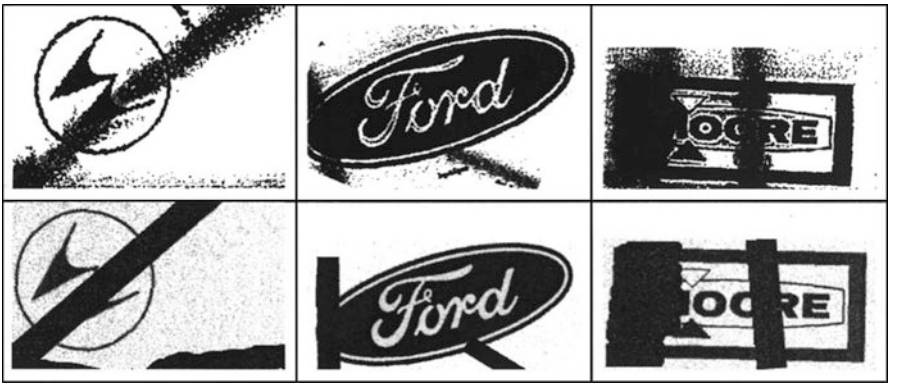

Fig. 18.8 Examples of noise produced by fax machines and Xerox copiers (*top*) and the corresponding images (*bottom*) obtained with noise models (Reproduced from [22])

Logo Detection and Removal in Images and Videos

This section reviews logo detection in natural images and video streams as well as the special topic of logo detection and removal in TV broadcasts. The background and processes involved in the applications discussed in this section are not fundamentally different from the applications discussed in the previous sections on "Trademark Retrieval Systems" and "Logo Recognition in Document Images." In particular, logo descriptors, logo recognition, and the principles of logo detection have already been introduced in previous sections. The main difference here is the distinct context of the application, namely, real-life images and videos, which in turn affects the efficiency and effectiveness of previously discussed approaches and guides the selection of the methodology. As a result, the focus of this section is on the particularities that the application context impinges on the methodological choices, and as such it is structured in an application-oriented manner.

First, a technical definition of the problem is given as well as an overview of typical applications and the needs they address. Subsequently, the detection of logos in images and video sequences is discussed and several methods are presented. Special attention is paid in methods regarding sports events which are the most common category of videos in which logos are searched for. This part finishes with a comparison table that summarizes the methodologies presented. The focus is then shifted on TV logo detection and removal where a comparison table is also provided. Finally, a summary of the open challenges in logo detection in images and videos is offered.

Problem Definition

Contrary to high-quality images of trademarks or scanned document images discussed before, this section focuses on the analysis of video sequences and

real-scene images that might or might not contain logos of interest. As will be appreciated next, although the main blocks of a logo detection system (logo descriptors, classification approaches, etc.) are similar to previously discussed ones, the methodologies developed for different applications such as trademark retrieval and document classification are generally not directly applicable in the domain of real-scene videos and images [8].

Before delving into the technical details, a description of the applications discussed in this section is offered. Similarly to text detection in videos and images (see ▸Chap. 25 (Text Localization and Recognition in Images and Video)), two variants of the problem are identified, depending on whether the logo of interest is part of the scene represented in the image or video sequence or it is a posteriori superimposed on it, a practice typically used with TV sequences.

Detection of Logos in Images and Videos

The appearance of logos in real scenes (videos or images) can be affected by various factors such as variations in illumination, occlusion, and projection distortions due to their placement in the scene. In addition to factors related to the scene layout, the media itself can introduce certain challenges and limitations, such as low resolution, blurring (especially in videos), and compression artifacts. On the other hand, as a matter of design, logos usually exhibit characteristic shapes and color schemes and are generally positioned so that they are easy to observe even in cluttered background [32].

The objective of a logo detection application is to detect the presence and location of a known logo in an unconstrained scene image or video. The logo to be detected is known to the application a priori, provided either as an image query or as a model learned off-line over various known instances of the logo. The number of logo models to be spotted is usually limited and rarely exceeds 20 or 30 models [8]. On the other hand, it is a usual requirement for the processing to take place in close to real time.

Typical applications of logo detection in videos and images are related to trademark use monitoring and advertisement impact assessment services. The objective of the former ones is to establish cases of logo misuse, while the latter focuses on quantifying the visibility achieved during specific events by measuring the time a company's logo was visible in the event media [8,37] (see, e.g., Fig. 18.9). A different branch of applications is built around mobile-based logo detection, where a typical goal is to provide the user with supplementary information about the corresponding company [44].

TV Logo Detection and Removal

Contrary to scene logos, logos superimposed on the video or images present a different set of challenges. The most emblematic application of superimposed logos is as a form of visible watermark, as they provide a direct assertion of content ownership. Superimposed logos are extensively used in the television broadcasting industry [19, 24, 59]. Almost all television broadcasts feature the television channel's logo in a prominent place, usually one of the video frame corners.

Fig. 18.9 Logos in sports videos (Reproduced from [37])

Superimposed logos can be opaque, semitransparent, or (partially) animated [82]. An opaquely superimposed logo overlaps a portion of the image content (see Fig. 18.10a), while a semitransparently superimposed logo blends with the host media (see Fig. 18.10b) allowing the original content to remain partially visible.

Fig. 18.10 TV logo types (**a**) opaque logo and (**b**) transparent logo (Reproduced from [89])

A typical challenge in terms of logo detection stems from semitransparent and animated logos which are not normative in terms of length and type of motion.

A frequent practice creating a different set of challenges is the superposition of multiple logos on the same content. This is typical of rebroadcast media, where the logo of the original station can be overlapped by the logo of the broadcasting one. TV logo recognition can also be beneficial for archival purposes in order to navigate in large video archives as well as for commercial detection since TV logos partly or fully disappear during commercials.

The main reason for detecting superimposed logos, apart from establishing the ownership of the media, is to automatically restore the substrate media. In such applications, once the logo is detected, it is removed and replaced by an estimation of the original media calculated from the surrounding visual content. This process is frequently called image inpainting, a term that is borrowed from the arts used to describe a restoration process for damaged paintings [54, 89].

Detection and Retrieval of Logos in Images and Videos

There are two main categories of algorithms which are widely used for logo detection in images and videos: methods based on color features and methods based on features at local keypoints. The methods based on color features are usually sliding window approaches that involve a classification step that compares the query logo image to subareas of the target image. A variety of color spaces are used in the literature, for example, RGB [14], HSV [60], and CIE-LAB [52], and most of the approaches use histogram representations; therefore, they can be efficiently implemented using integral histogram approaches. On the other hand, local keypoints-based methods use point correspondences for matching object templates involving a voting scheme. They regard an object as a set of local keypoints and fit an affine transformation to simulate their geometric consistency. Both are valid approaches and both have been proven to provide reasonable results under certain conditions.

Sliding Window-Based Methods

Logo detection and localization in color-based approaches is usually performed using a sliding windows framework [14, 38, 52, 60]. In order to reduce computation time, the target image is often subsampled before further processing [38, 52, 60]. For detecting multiple sizes of the input query, multiple scale factors of the input query are also considered. Choosing equally spaced scale factors is one option, but proper size matching is more important at smaller sizes than at larger ones [52]. Regarding similarity measure, histogram intersection is commonly adopted for comparison purposes.

Color is one of the most discriminatory and commonly used features that are frequently employed for logo detection tasks in unconstrained images. For example, Chang and Krumm [14] proposed an object detection algorithm using color co-occurrence histograms (CCH) as an object representation. They quantized multiple object models to a small number of colors using a k-means clustering algorithm in the RGB color space and then quantized the test images using the same color clusters.

However, color clustering methods are sensitive to changes in illumination and object size which can vary widely between images. The robustness of such approaches can be significantly increased by considering spatial relationships between the colors while allowing for some degree of distortion (see also logo representations combining topological and color features as discussed in the context of trademark retrieval). In this line, Luo and Crandall [52] proposed a method that captures the separation of pairs of color pixels at different spatial distances when these pixels lie in edge neighborhoods, overcoming the problem with the disproportionate energy contributions demonstrated by a single color on the CCH. A pixel is considered an edge pixel if it has different color from any of its eight connected neighbors. This implies that only pixels that are on, or very close to, an edge are included and these tend to be pixels containing the most important spatial–color information. The notion of edge pixels has been further explored by Phan and Androutsos [60] who created an edge map of valid edge points by using vector order statistics that depend on edge gradients. In another approach [38] wavelet decomposition coefficients are introduced and a co-occurrence histogram of both color and wavelet directional detail information is adopted.

An important decision in approaches based on color information is the color space in which color differences are interpreted. The quantization scheme or distance function must be carefully designed to ensure that perceptually similar colors are spatially close and mapped to the same quantized color value. The RGB color space is not perceptually uniform prompting researchers to use more perceptually uniform color spaces and color appearance models. For example, in [52] the CIE LAB color space in conjunction with the ISCC–NBS system is adopted, while in [60] the colors are classified under similar hue orientations in the HSV color space.

Overall, employing color for logo detection in real scenes is a natural choice since color information is a dominant feature in this type of medium. Having said

that, it is important to note that color is not necessarily a good feature for all logos. For certain logos (e.g., the Starbucks or the Coca-Cola ones) the color scheme is a registered feature of the trademark and reproduced faithfully in every instance of the logo. But there is a huge number of logos for which the color scheme is not that important and logos might be reproduced in arbitrary color combinations (e.g., the Nike or the Apple logo). For these logos, purely color-based descriptors might not be the best option.

Color histogram-based methods can be efficiently implemented based on techniques such as integral images or dynamic programming. On the other hand, finding logos in different scales requires the application of the method in several window sizes which results to a very large number of patches that are too expensive to search exhaustively, even in small-sized target images. Furthermore, sliding windows provide poor results when the objects in the target image, are occluded or are subject to heavy deformations.

Keypoint Correspondence-Based Methods

Keypoint-based local feature methods have also become a common choice regarding logo detection in natural images through keypoints correspondences [44, 46]. They have been successfully used to describe logos and obtain flexible matching techniques that are robust to partial occlusions as well as linear and nonlinear geometric transformations. In these methods a voting scheme is usually involved in order to detect and localize potential logo instances in the target image. Typically, an affine transformation is then calculated that maps the logo onto the target image and patches which accumulate above a certain threshold of votes are retained.

Besides the voting approach, several authors address the problem of properly grouping the matched local descriptors in the target image in order to detect potential logo instances. For instance, Kleban et al. [46] enrolled data mining association rules that capture frequent spatial configurations of quantized local SIFT descriptors. It is an extension of an idea introduced by Quack et al. [62] where association rules are employed to select features for object detection. The resulting different types of rules are appropriately weighted, and logo localization is performed by grouping selected features using mutual rule matches with representative training examples.

Spectral saliency analysis is another tool to localize logo instances by detecting the local high-frequency regions of an image. It is based on the hypothesis that a natural image consists of two parts, the redundant part and the novel part. However, it may result to unwanted parts with similar local frequency characteristics. To overcome this, Gao et al. [32] propose a partial spatial context descriptor to formulate the spatial distribution for the set of matching SURF keypoints [9]. The descriptor is based on a bag-of-words formalism and describes both the transformation consistency and contextual information of the point set. In a different line, Meng et al. [55] detect logo instances by finding subimages where the largest point-wise mutual information towards the query is measured. A branch-and-bound

search ranks the subimages by the mutual information, and a relevance feedback scheme further improves the results by verifying relevant and irrelevant subimages.

A common way to calculate the geometric consistency of a potential logo instance is by estimating an affine transformation model between the query and the point set. For this purpose the RANSAC algorithm [31] is often employed in order to model the transformation, resulting to a number of inliers, that is, matches that fit the model. However, using a fixed threshold for the number of inliers as a spatial verification criterion has many drawbacks since it depends on many factors like the average number of keypoints in the query and the image, the redundancy of keypoints, and the size of the query. To overcome this, Joly and Buisson [44] proposed an "a contrario" adaptive thresholding approach, adapted to geometric consistency scores. A global geometric consistency is achieved by applying a threshold in normalized scores where the scoring of matches is directly related to the statistical dependence between the spatial positions of the query and the matched keypoints.

In general, methods based on keypoints are usually faster than the ones based on color. They are also adequate in cases where robustness to occlusions and clutter is required. A limitation of these methods is that in a typical application, thousands of nonrelevant keypoints are produced in the target image, where nonrelevant refers to keypoints not belonging to any logo region. This leads to increased storage requirements, while despite the large number of candidate keypoints, the annotated logos may still contain only 0, 1, or 2 points, which is insufficient for recognition.

Local descriptors are based on intensity values ignoring the color information which is a prominent logo feature in many cases. To increase illumination invariance and discriminative power, several color descriptors have been proposed in the literature. In a recent study, Van de Sande et al. [79] provide a thorough discussion on the invariance properties and the distinctiveness of color descriptors. In a logo detection perspective, they can be seen as an attempt to bridge the gap between color-and intensity-based methodologies.

Logo Detection in Videos

Similar to still images, color information is a key feature frequently used in the literature regarding logo detection in video sequences [4, 20, 37]. Invariance under color distortions is an important task for these methods due to the constantly changing color information in the context of video streams. For instance, in [37] the effects of illumination intensity (e.g., due to clouds or other environmental conditions) are reduced by using only the chrominance components of the luminance–chrominance space. Alternatively, the photometric invariant color space [33] has been also opted for since it provides robustness to changes of surface orientation, illumination direction, and intensity [4].

The local color distribution can also be used as an indicator of areas of interest in video frames. For example, Hollander and Hanjalic [20] detect potential logo areas by searching for homogeneously colored regions that are surrounded by large color differences. The image area is represented by intensity profiles along lines, and logo recognition is performed by matching these lines with the line profiles of

the query logo models. Low-cost color processing is also involved in [37] to detect candidate logo regions that are then recognized using scale-normalized Gaussian receptive fields computed over a limited region of interest. In this manner, a model of the logo's visual appearance can be created based on a small number of sample query images. The identification of the logo instances can be based either on the distance between histograms or on probabilistic measures.

Logo detection can be further assisted by closely related techniques that focus on the detection of broader image regions that may enclose the logo under consideration. For example, billboards in sports broadcasting streams are located on the side of the field and are usually the place where brand logos are placed. Therefore, knowing the exact position of billboards significantly narrows down the spatial search space for logo detection. On the other hand, developing a billboard analysis and detection system faces several challenges on its own right [4]. Watve and Sural [83] proposed a method where frames within video shots are segmented to locate possible regions of interests in a frame where billboards are potentially present. A combination of local and global features is also employed for detecting individual billboards by matching them with a set of given templates.

Local keypoints associated with local descriptors are the basis for several methods regarding logo detection on videos. The logo queries are matched against the content extracted from every frame of the video to compute a "match score" indicating the likelihood that a particular logo occurs at any given point in the video [8]. It is evident that the requirement for near real-time response poses a computational burden to the detection of logos in video sequences in a frame-by-frame matching basis. Depending on the application, a common approach is to separate the detection process into an off-line preprocessing step and an online querying step. The benefit of this approach is that matches are effectively precomputed so that at run-time frames and shots containing any particular object can be retrieved with no delay. An example of such a system is given by Sivic and Zisserman [75] where an object retrieval approach is presented which incorporates text retrieval techniques in order to search and localize all the occurrences of an object in a video. The query is provided by a user-specified subpart of an image, and a bag-of-words model is applied in the visual domain by generating a codebook of affine covariant features, represented as SIFT descriptors. In terms of computational efficiency, their method demonstrates two orders of magnitude speedup without significant loss in performance when compared to the standard frame-to-frame matching. More interestingly, the system could efficiently search for instances of an "unknown" logo that was not part of the video stream, that is, its descriptors have not been pre-calculated during the off-line step.

Ensuring the temporal continuity of logo instances or candidate regions within a video shot facilitates the detection process. In this line, in [75] the regions detected in each frame of the video are tracked using a simple constant velocity dynamical model and correlation. Any region which does not survive for more than three frames is rejected. On the other hand, a dedicated tracking module is involved in [37] that uses Kalman filtering. It is based on the estimated position and size of the current logo in order to predict a region of interest in the next frame.

Concluding, it should be noticed that not all of the techniques presented so far are fully automated but instead some of them rely somehow on human intervention. For example, in [8] the time intervals of the video likely to contain the logos are used to drive a user interface through which a human annotator is involved in order to validate the automatic annotation results. Similarly, in [37] a supervisor coordinates the different modules, keeps track of targets that have been identified, and halts the tracking of a target region when identification fails. Table 18.5 provides a comparative summary that summarizes the aforementioned methods for logo detection in images and videos.

TV Logo Detection and Removal

This section focuses on TV logo detection in broadcast videos which poses a different problem than the one addressed in the last section. Several methods are presented that tackle the problem of finding superimposed logos that may be opaque, transparent, or even animated usually surrounded by constantly changing background content. The task of removing the TV logo is also discussed focusing on the smooth replacement of the removed logo by properly reconstructed background content.

Logo Detection

TV Logo detection is meant to locate and extract a logo within a sequence of images. Broadcasters' logos are typically over imposed inside one of the four corners of the video frame [3, 59]. Taking this fact into consideration, a significant acceleration of the overall processing time can be achieved, which is essential for real-time implementation. Thus, it can be assumed that the probability of the logos appearing in the four corners of the video frame is higher than the probability of its appearing in the center [89]. Alternatively, prior information can be involved by building a table that contains information regarding in which of the four corners of the video frame the logo under consideration may appear [69]. A general approach that is applied to detect both opaque and semitransparent logos is to find areas in the image that provide some kind of spatiotemporal persistency in terms of properly selected features. For example, in [59] time-averaged edge fields are employed based on classical edge detection methods while preserving the persistency of the extracted edges over a number of frames. In a similar way, Albiol et al. [3] seek for areas with stable contours and they use time-averaged gradients in order to detect them.

A different variety of methods employ pixel-wise approaches which are based on the observation that image areas where logo is overlaid show luminance variance values in a narrower interval than the rest of the image areas, depending on the logo transparency [54, 69, 89]. For example, in opaque logos it is assumed that the video content changes over time except for the logo area and any differences between subsequent frames at the logo position occur due to noise. In general, the objective of the temporal segmentation is to find minimal luminance variance regions in every

Table 18.5 Comparative summary of image and video logo detection and retrieval methods

Name	Year	Features + descriptors	Detection + retrieval	Medium	Dataset
Hollander and Handjalic [20]	2003	Affine invariant binary color profiles	String matching	Video	5 logos in 5 sequences
Sivic and Zisserman [75]	2003	Vector-quantized SIFT descriptors	Bag-of-features matching + tf–idf weighted frequency vectors + spatial consistency voting	Video	6 objects in 2 full-length films
Hall et al. [37]	2004	Multidimensional color histograms	Scale-normalized receptive fields + Kalman-filtering tracking	Video	3 logos in 2 formula one videos
Luo and Crandall [52]	2006	Color edge co-occurrence histogram	Spatial-color joint probability functions + normalized cross correlation	Image	Several images from the Web
Bagdanov et al. [8]	2007	SIFT descriptors	Bag-of-features matching + normalized match score	Video	6 logos in 3 sports videos
Hesson and Androutsos [38]	2008	Wavelet decomposition coefficients + color	Sliding non-overlapping windows of varying size in multiple scales + histogram intersection	Image	2 query logos + 3,000 test images
Kleban et al. [46]	2008	Keypoints by Harris affine detector + SIFT descriptors	Spatial pyramid mining of association rules + DBSCAN density-based clustering	Image	7 query logos in 974 images from the Web
Phan and Androutsos [10]	2009	Color edge cooccurrence histogram + vector order statistics + color quantization method based on HSV color space	Sliding non-overlapping windows in multiple scales + histogram intersection	Image	400 logo images + 5,000 unrelated images.
Gao et al. [32]	2009	Spectral + spatial saliency analysis, SURF features	Maximum saliency density, similarity based on partial spatial context	Image	10 query logos + 10,016 test images from the Web

(continued)

Table 18.5 (continued)

Name	Year	Features + descriptors	Detection + retrieval	Medium	Dataset
Joly and Buisson [44]	2009	SIFT descriptors	Multi-Probe Locality Sensitive Hashing + RANSAC on affine transformation model + "a contrario" query expansion	Image	Oxford Building dataset + BelgaLogos
Meng et al. [55]	2010	Keypoints by Harris affine detector + SIFT descriptors	Branch and bound search + maximum mutual information + relevance feedback	Image	BelgaLogos

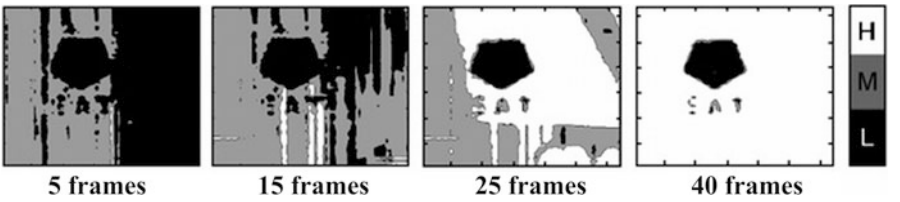

Fig. 18.11 Evolution of the high (*H*), medium (*M*), and low (*L*) luminance variance areas of the LVI as the number of processed frames grows (Reproduced from [19])

frame. In this line, Cozar et al. [19] describe a method that involves the luminance variance image (LVI). The LVI is an image with the same size as the video frames in which a pixel is represented by the difference between the maximum and minimum luminance values in the corresponding pixel location along the sequence of frames. The minimal luminance variance regions can be extracted by applying a properly defined threshold. Figure 18.11 depicts an example where high, medium, and low luminance variance areas of the LVI are shown as the number of processed frames grows. Clearly, such approaches are affected by the set of frames used for variance calculation and the value of the threshold.

Instead of using the logo as one entity, Yan et al. [89] introduced the notion of logolets, which are small image regions corresponding to logo parts. A Bayesian classifier framework is then used in combination with an ANN trained to detect the logolets in video frames. Duffner and Garcia [24] explored further the use of ANN by proposing a system for transparent logo detection where the raw input image is treated as a whole without any assumptions about the features to extract or the areas to be analyzed. For this purpose, a multi-scale convolutional ANN architecture is utilized that derives problem-specific feature extractors from a large training set of logo and non-logo patterns.

In case of animated logos, the approach to be followed depends significantly on the percentage of the logo area that is changing through time. For instance,

partially animated logos can be regarded as opaque ones, because they can be detected through their immovable parts. However, in the general case, a more delicate treatment is required since the temporal persistency has to be preserved not just for a single frame instance but for a set of consecutive frames that form the complete logo animation circle. In this line, Wang et al. [82] use gradient information as low-level visual features and propose a template-matching approach that seeks persistent gradients over multiple video frames. In [29] a method tailored specifically for animated logos is proposed that utilizes a multi-frame contour representation which is matched by the contours of a test video using a voting-based decision scheme.

Typically, the result of the detection methods is a binary mask that indicates the region where the TV logo is placed. The images that denote the difference between subsequent frames are binarized by thresholding in order to obtain initial logo masks, and the final logo mask is obtained, for example, by contour relaxation based on Markov Random Fields [54]. It is also common to further refine the logo mask by morphological processing in order to reduce spurious pixels and fill holes [59]. Additional constraints may also be applied regarding the shape or the size of the mask in order to eliminate improbable results [3].

Overall, TV logo detection is a process considering both spatial and temporal domains. Introducing a spatial persistency requirement benefits the detection process. This requirement is derived from the observation that no change occurs over time for the TV logo position and that the logo appears in one of the four corners of the video frame (or, more rarely, interchangeable in different corners depending on the video stream content, e.g., commercials). On the other hand, the persistency of the logo over time is closely related to the number of subsequent frames under consideration. A low number of frames is not sufficient to eliminate noise from accidentally stable background content that could increase false alarms. On the contrary, too many frames tend to blur a logo's inherent features that could increase missed segments [82].

Logo Removal

The efficient detection of the logo mask is an important prerequisite for automatic TV logo removal in order to improve the viewing experience of rebroadcast programs or to replace the existing logo with another one [54, 82]. An intuitive approach is to find the best replacement patch for the removed region within the video shot. Indeed, if the motion of the logo region is sufficient enough, then the region underneath the logo may be exposed in a nearby frame. In this case, matching has to be based on a properly selected measure that demonstrates low sensitivity to small perturbations of the image while being tolerant in small position errors. Yan et al. [89] follow this approach in order to remove small logos by adopting the Hausdorff distance. However, if the logo region is too large, the overlapping will result in observable edges for the video region. In these cases a video inpainting approach is usually opted for.

Video inpainting can be considered as the approximation of the covered region by using the surrounding information and their differences. However, two main issues arise with this technique, namely, the size of the logo area and how to

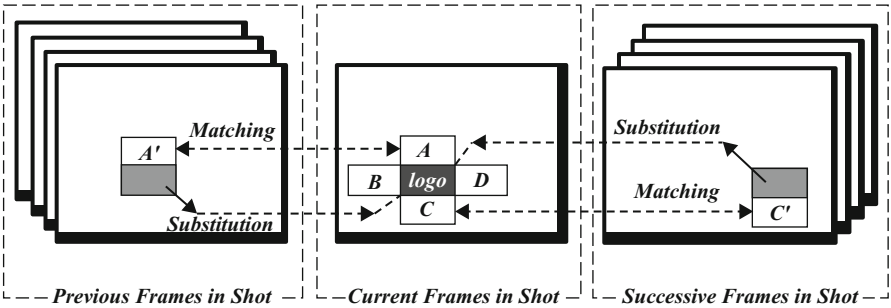

Fig. 18.12 Matching-based algorithm for region overlapping (Reproduced from [89])

maintain coherence between adjacent video frames that is distorted by unwanted edges appearing due to the overlapping processing. In [89] an extension of the 2D gradients in the image inpainting technique is proposed which uses 3D gradients exploiting the temporal correlations in video. Figure 18.12 shows an example where in order to obtain the substitution region of the logo, the regions in the previous and next frame are considered, with the most similar pair among all the candidate pairs being selected. The corresponding region is then used to replace the logo region.

Video inpainting is also used by Wang et al. [82] where a multivalued image regularization with Partial Differential Equations (PDE) is proposed. PDE-based regularization may be seen as the local smoothing of an image along defined directions depending themselves on the local configuration of the pixel intensities [78]. The target is to smooth the image while preserving its edges (discontinuities in image intensities), i.e., perform a local smoothing mostly along directions of the edges, avoiding smoothing orthogonally to these edges. In [82] PDE is employed for inpainting the logo regions while preserving the local geometry of the multivalued image discontinuities. The inpainting of the logo region with the neighborhood area requires a temporal set of frames in order to compute a structure tensor in the spatiotemporal domain.

TV logo removal can also be treated as an extrapolation problem. In this case the image area surrounding the logo is extrapolated using a frequency-selective method and used to replace the logo. Linear combinations of weighted basis functions can be used as parametric models in order to approximate the support area of a logo. In this context, spectral estimation methods are involved in order to describe the logo area by a few dominant features while the logo is given by extrapolation. For example, Meisinger et al. [54] use two-dimensional Discrete Fourier Transform basis functions since they are especially suited in order to conceal monotone areas, edges, and noise-like regions. In case of large logos, the logo is partitioned into blocks and the different blocks are processed subsequently. For inpainting the logo, first the corner blocks and then the inner blocks are inpainted, in order to exploit as much surrounding data as possible.

In general, video inpainting is a powerful tool that delivers visually pleasant results. However, the current methods still encounter difficulties, for example, when dynamic-texture backgrounds are under consideration or in case where occluded objects are involved in the video scene. In such cases the spatiotemporal inconsistency leads to visually annoying artifacts and further refinements are required. For example, in [89] graph cut textures are used to find a similar block and blend the visible edges that arise due to overlapping.

A summary of the key aspects of the above methods regarding TV logo detection and removal methods is provided in Table 18.6.

Open Challenges in Logo Detection and Removal in Images and Videos

Generalization

Searching for logos in databases of unconstrained color images is a time-consuming and labor-intensive task since the applicability of any method is subject to many uncontrollable factors such as lighting, deformation, and occlusion. Video streams are further characterized by perspective deformation due to pan, tilt, and zooming operations of the camera, motion blur, as well as often rapid changes in the illumination conditions. Several methods address these issues by requiring a large number of sample queries acquired under various conditions, while others are based on models that aim to predict a large number of potentially problematic cases. However, there is still a lack of approaches that will expose robustness and generalization for logo detection in a variety of conditions.

Similarly, in case of videos the detection efficiency is heavily related to the variance that characterizes the context in video sequences as well as on the representativeness of the query logo model. It has been reported that synthetic query logos perform worse in terms of precision and recall than logo instances cropped directly from the video [8]. This is due to the fact that many other logos consisting of mostly text and graphics are confused for the synthetic logo models. Using more than one query logo model is a solution but again the sample logos should cover reliably the variations during the video. In [37] it is argued that 6–8 instances are sufficient, but in any case the color model for each logo must be acquired from images under actual illumination conditions to reduce the effects of illumination changes.

Features Applicability

While color-based approaches have demonstrated reliable results in logo detection tasks, it is evident that raw color information might be quite instable due to illumination and shadows. There are methodologies for illumination estimation and shadow removal, but of course including them in the pipeline would imply a considerable time cost. On the other hand, even if some kind of color normalization is assumed, then there are still many alternative human-oriented descriptions for color features that yield better results (see, e.g., the color naming discussion in

Table 18.6 Comparative summary of TV logo detection and removal methods

Name	Year	Detection	Localization + tracking	Logo removal	Logo type	Dataset
Albiol et al. [3]	2004	Stable contours	Temporal gradient analysis + morphological operations	N/A	Opaque, transparent	6 TV channels
Yan et al. [89]	2005	Frame differentiation + logolets	Bayesian classifier framework combination with an ANN	Overlapping matching + video inpainting by 3D gradients + graph cut textures	Opaque	Video clips collected from various Web sites
Meisinger et al. [54]	2005	Frame differentiation	Binarization through thresholding + contour relaxation based on Markov Random Fields	Extrapolation of surrounding image by frequency-selective method	Opaque, semitransparent	TV streams
Wang et al. [82]	2006	Multispectral gradient information	Generalized interframe gradient + adaptive thresholding	PDE-based regularization	Opaque, transparent, animated	4 h from TRECVID'05 + several TV channels
Duffer and Garcia [24]	2006	Train ANN with positive and negative examples	Subsampling in pyramid of images + localization by proximity in image and scale space	N/A	Transparent	800/257 images from TV with/without logo

Santos and Kim [69]	2006	Edge detection by magnitude of gradient + time averaging	Cross correlation + hypothesis testing	N/A	Opaque, semitransparent, partially animated	24 h from 10 channels
Cozar et al. [19]	2007	Minimal luminance variance region	Spatiotemporal segmentation + nonlinear diffusion filters	N/A	Opaque, semitransparent	6 h of 45 broadcasted video sequences from different TV channels
Esen et al. [29]	2008	Multiframe contour representation	Compare to sample video by Generalized Hough transform + majority voting	N/A	Animated	27 h of 3 channels
Ozay and Sankur [59]	2009	Canny edge detection + time-averaged edges + SVM for best feature selection	Binarization hysteresis threshold + morphological post processing	N/A	Opaque, transparent	240 video samples from 12 TV channels

the section on "Trademark Retrieval Systems"). This would make sense as logos comprise a limited number of named colors.

In the case of local keypoints, spatial consistency is another important aspect of the matching process since it helps filtering out false matches. It can be measured either by simply requiring that neighboring matches in the query lie in a surrounding area in the video frame or more strictly by requiring the same spatial layout in the neighboring matches of both the query image and the video frame [75]. Another problem is posed by logo models that give rise to a relatively low number of feature points, causing the matching process to become unreliable.

Finally, despite that several robust features have been extensively used (e.g., color, shape, local keypoints), contextual information is rarely explored although it may provide a robust estimate of the probability regarding the logo's presence, position, and scale [58]. Indeed, global scene representations based on aggregated statistics of low-level features may act as sources of contextual information that can improve the detection efficiency while providing tolerance to image degradation and illumination changes.

Scalability

Logo detection in large image databases is a time-consuming process, with the unconstrained nature of real-scene images (as opposed, e.g., to the document images discussed before) creating a demand for substantially more complex analysis methodologies.

The computational load of processing a typical video stream, comprising tens of thousands of frames, is heavy, even if approaches such as frame sampling are applied. As a result, real-time requirements are hardly met by state of the art methods. An even more important observation though is that the majority of existing logo detection methodologies would be difficult to scale gracefully with the number of frames processed or the number of logo models queried. This is particularly true for methodologies relying on indexing of large numbers of local features and analyzing correspondences between them. Having said that, there is a lot of promising research in the field and a lot of investment on big-data processing, thus there are expectations that new breakthroughs will be achieved in the near future.

Conclusion

This chapter discussed various applications related to the detection, localization, and recognition of logos and trademarks in different media. The chapter attempted to cover the considerable breadth and depth of the topic at hand by using three identified applications as the guide for the discussion.

Starting with the application of trademark retrieval systems, typical trademark representations and descriptors were introduced, while aspects related to perceptual organization of graphical entities were discussed as well as typical practices in indexing and retrieval. Subsequently, in the context of logo recognition in document

images, the use of logos for classification was examined and the application of local keypoints-based descriptors and logo spotting was introduced. Finally, the application of logo detection and removal in images and videos was studied, focusing on the challenges that real-scene media present as well as the typical application of logo removal and restoration of the original substrate in TV broadcasts. In all cases, summary tables of published work were provided, linking to the different aspects of the problem discussed in the text.

As the range of applications examined is extensive and diverse, it is difficult, if not meaningless, to compile a short list of generic open challenges, inviting instead the interested reader to review the respective "open challenges" sections for each application discussed in this chapter. Nevertheless, it is worth noting that one open problem repeatedly encountered in different applications is the issue of scalability of the proposed methodologies, so much in terms of the number of trademark models treated as in terms of the processing time required.

In summary, logo and trademark recognition is a multifaceted area of research, touching upon different domains and covering a variety of distinct real-world applications. As such, it is an interesting topic for multidisciplinary research and an attractive evaluation setting for a wide range of disciplines and topics including document analysis, indexing and retrieval, perceptual organization, and multimodal data fusion to mention just a few.

Description of Reference Datasets

A good source for datasets and tools are the Web sites of IAPR TC10 ("Graphics Recognition") and TC11 ("Reading Systems")[6] that host or link to the latest available information. The information given below was correct at the time of writing, but the readers are encouraged to look at the above sites for most up-to-date information. See also ▸Chaps. 29 (Datasets and Annotations for Document Analysis and Recognition) and ▸30 (Tools and Metrics for Document Analysis Systems Evaluation) of this book for more generic tools and datasets.

Trademark Retrieval Systems

Public datasets of trademarks with associated ground truth information are difficult to encounter, although individual trademarks are searchable and downloadable one by one through the Web sites of trademark registration offices. The interested readers can check, for example, the CTM-ONLINE service of the OHIM,[7] the "Localizador de Marcas" of the Spanish Patent and Trademark Office,[8] or the "Trademark Electronic Search System" of the US Patent and Trademark Office.[9]

In terms of collections of trademarks, the most used trademarks dataset is the "MPEG-7 Still Images Content Set," Item S8, which comprises about 3,000

B&W geometric trademark images captured by a scanner, provided by the Korean Industrial Property Office more than a decade ago [12]. The dataset is not publicly available, and according to its licensing agreement, its use is strictly for noncommercial purposes.

An alternative collection of trademarks and logos provided by the IBM Almaden Research Center is also available at the time of writing[10] [40, 41], although the licensing particulars do not seem to be clear.

Logo-Based Document Classification

The evaluation in most of the methods regarding document classification based on logos is performed using the Tobacco-800 dataset [76] and the University of Maryland logo database [49]. Tobacco-800 is a public subset of the IIT CDIP Test Collection based on 42 million pages of documents (in 7 million multipage TIFF images) obtained from UCSF and released by tobacco companies under the Master Settlement Agreement. The documents were collected and scanned using a wide variety of equipment over time providing therefore a comprehensive dataset characterized by large variability. The University of Maryland logo database is another commonly used dataset that contains 106 distinct logo images in TIFF format. The logos in the dataset have a variety of sizes ranging from 134×116 pixels up to 629×671 pixels. Some examples of both datasets are shown in Fig. 18.13.

Logo Detection and Removal in Images and Videos

BelgaLogos [44] is a natural image collection specifically created for logo detection and retrieval evaluation. The dataset is freely available for research purpose only. It considers 26 different logos and consists of 10,000 images covering diverse categories of objects and events. A given image can contain one or several logos or no logo at all. On the other hand, the World Wide Web is another popular source for unconstrained images; therefore, several authors tend to compile their own private datasets which typically consist of several hundred images wherein a small number of logos is searched for.

In the case of videos, logos are usually part of billboards, banners, and other physical advertising media. Most commonly, such advertisements appear in sports broadcasting like soccer and football fields, formula one race circuits, and tennis courts. Therefore, the majority of methods in the literature use sports videos as evaluation datasets. Similarly, for TV logo detection and removal evaluation, a diversity of broadcast video and TV channels are used. However, there is still a lack of any standardized datasets, while it is clear that such data associated with ground truth data would be a valuable aid for evaluation purposes.

Fig. 18.13 Example logos from (**a**) Tobacco-800 dataset and (**b**) University of Maryland logo database

Cross-References

▸An Overview of Symbol Recognition
▸Datasets and Annotations for Document Analysis and Recognition
▸Graphics Recognition Techniques
▸Image Based Retrieval and Keyword Spotting in Documents
▸Page Similarity and Classification
▸Text Localization and Recognition in Images and Video
▸Tools and Metrics for Document Analysis Systems Evaluation

Notes

[1] United States Patent and Trademark Office, "Performance and Accountability Report Fiscal Year 2011."

[2] Source "OHMI Manual of Trade Mark Practice" Part 2, Chapter 2C, OHMI, accessible online: http://oami.europa.eu/ows/rw/pages/CTM/legalReferences/guidelines/OHIMManual.en.do

[3] Note that there are also "color marks" per se, referring to the situation that a particular color without any specified contours is registered as a trademark. This type of trademarks is out of the context of this chapter.

[4] OHMI, "Manual of Trade Mark Practice," online resource, accessed on May 2012, Source: http://oami.europa.eu/ows/rw/pages/CTM/legalReferences/guidelines/OHIMManual.en.do

[5] Source: "Implementing a Digital Mailroom," White Paper, Datafinity Ltd, June 2009, UK.

[6] http://www.iapr-tc10.org/ and http://www.iapr-tc11.org/ respectively.

[7] http://oami.europa.eu/ows/rw/pages/QPLUS/databases/searchCTM.en.do (last accessed on Sept. 2012).

[8] http://sitadex.oepm.es/Localizador/homeLocalizador.jsp (last accessed on Sept. 2012).

[9] http://www.uspto.gov/trademarks/index.jsp (last accessed on Sept. 2012).

[10] http://www.eurecom.fr/~huet/work.html (last accessed on Sept. 2012).

References

1. Alajlan N (2007) Retrieval of hand-sketched envelopes in logo images. Lect Notes Comput Sci 4633/2007:436–446
2. Alajlan N, Kamel MS, Freeman G (2006) Multi-object image retrieval based on shape and topology. Signal Process Image Commun 21(10):904–918
3. Albiol A, Fulla MJ, Albiol A, Torres L (2004) Detection of TV commercials. In: Proceedings of the international conference on acoustics, speech and signal processing, Montreal, pp 541–544
4. Aldershoff F, Gevers T (2004) Visual tracking and localisation of billboards in streamed soccer matches. SPIE Electron Imaging 5307:408–416
5. Alwis S, Austin J (1999) Trademark image retrieval using multiple features. In: Proceedings of the challenge of image retrieval (CIR-99), BCS electronic workshops in computing, Newcastle-upon-Tyne
6. Amir A, Lindenbaum M (1998) A generic grouping algorithm and its quantitative analysis. IEEE Trans Pattern Anal Mach Intell 20(2):168–185
7. Baeza-Yates R, Ribeiro-Neta B (2011) Modern information retrieval: the concepts and technology behind search, 2nd edn. Addison-Wesley, New York
8. Bagdanov AD, Ballan L, Bertini M, Bimbo AD (2007) Trademark matching and retrieval in sports video databases, in James Ze Wang; Nozha Boujemaa; Alberto Del Bimbo & Jia Li, ed., Multimedia Information Retrieval, ACM, New York, pp 79–86
9. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: ECCV, Graz
10. Belkin NJ, Kantor P, Fox EA, Shaw JA (1995) Combining evidence of multiple query representations for information retrieval. Inf Process Manag 31(3):431–448
11. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(24):509–522
12. Bober M, Preteux F, Kim Y-M (2001) MPEG-7 visual shape descriptors. Technical report, Ref: VIL01-D112, Mitsubishi Electric
13. Chan DY-M, King I (1999) Genetic algorithm for weights assignment in dissimilarity for trademark retrieval. In: Huijsmans DP, Smeulders AWM (eds) VISUAL'99, Amsterdam. LNCS 1614, pp 557–565

14. Chang P, Krumm J (1999) Object recognition with color cooccurrence histograms. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Fort Collins, pp 498–504
15. Chen J, Leung MK, Gao Y (2003) Noisy logo recognition using line segment Hausdorff distance. Pattern Recognit 36(4):943–955
16. Chen J, Wang L, Chen D (2011) Logo recognition: theory and practice. CRC, Boca Raton
17. Chou T-C, Cheng S-C (2006) Design and implementation of a semantic image classification and retrieval of organizational memory information systems using analytical hierarchy process. Omega 34:125–134. Elsevier
18. Ciocca G, Schettini R (2001) Content-based similarity retrieval of trademarks using relevance feedback. Pattern Recognit 34:1639–1655
19. Cózar JR, Guil N, González-Linares JM, Zapata EL, Izquierdo E (2007) Logotype detection to support semantic-based video annotation. Signal Process Image Commun 22(7–8):669–679
20. den Hollander RJM, Hanjalic A (2003) Logo recognition in video stills by string matching. In: Proceedings of IEEE international conference on image processing (ICIP), Barcelona, pp 517–520
21. Desolneux A, Moisan L, Morel J-M (2008) From Gestalt theory to image analysis: a probabilistic approach. Springer, New York
22. Diligenti M, Gori M, Maggini M, Martinelli E (2001) Adaptive graphical pattern recognition for the classification of company logos. Pattern Recognit 34:2049–2061
23. Doermann D, Rivlin E, Weiss I (1996) Applying algebraic and differential invariants for logo recognition. Mach Vis Appl 9(2):73–86
24. Duffner S, Garcia C (2006) A neural scheme for robust detection of transparent logos in TV programs. Lecture notes in computer science—II, vol 4132. Springer, Berlin, pp 14–23
25. Eakins JP, Shields K, Boardman JM (1996) Artisan—a shape retrieval system based on boundary family indexing. In: Sethi IK, Jain RC (eds) Storage and retrieval for image and video databases IV (Proc SPIE 2670). SPIE, Bellingham, pp 17–28
26. Eakins JP, Boardman JM, Graham ME (1998) Similarity retrieval of trademark images. IEEE Multimed 5(2):53–63
27. Eakins JP, Riley KJ, Edwards JD (2003) Shape feature matching for trademark image retrieval. In: Bakker EM et al (eds) CIVR 2003, Urbana-Champaign. LNCS 2728, pp 28–38
28. El Badawy O, Kamel M (2002) Shape-based image retrieval applied to trademark images. Int J Image Graph 2(3):375–393. World Scientific
29. Esen E, Soysal M, Ates TK, Saracoglu A, Aydin Alatan A (2008) A fast method for animated TV logo detection. In: CBMI, London, pp 236–241, June 2008
30. Fall CJ, Giraud-Carrier C (2005) Searching trademark databases for verbal similarities. World Pat Inf 27:135–143
31. Fischler M, Bolles R (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. CACM 24(6):381–395
32. Gao K, Lin S, Zhang Y, Tang S, Zhang D (2009) Logo detection based on spatial-spectral saliency and partial spatial context. In: Proceedings of the ICME, New York, pp 322–329
33. Gevers T, Stokman H (2004) Robust histogram construction from color invariants for object recognition. Trans Pattern Anal Mach Intell 24:113–118
34. Giacinto G, Roli F (2005) Instance-based relevance feedback for image retrieval. In: Saul LK, Weiss Y, Bottou L (eds) Advances in neural information processing systems, vol 17. MIT Press, Cambridge, MA, pp 489—496
35. Gori M, Maggini M, Marinai S, Sheng J, Soda G (2003) Edge-Backpropagation for noisy logo recognition. Pattern Recognit 36(1):103–110
36. Grossman DA, Frieder O (2004) Information retrieval: algorithms and heuristics, 2nd edn. Springer, Dordrecht
37. Hall D, Pelisson F, Riff O, Crowley JL (2004) Brand identification using Gaussian derivative histograms. Mach Vis Appl 16(1):41–46
38. Hesson A, Androustos D (2008) 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, Las Vegas, pp 1233–1236

39. Hsieh I-S, Fan K-C (2001) Multiple classifiers for color flag and trademark image retrieval. IEEE Trans Image Process 10(6):950
40. Huet B, Hancock ER (1999) Line pattern retrieval using relational histograms. IEEE Trans Pattern Anal Mach Intell 21(12):1363–1370
41. Huet B, Hancock ER (2002) Relational object recognition from large structural libraries. Pattern Recognit 35(9):1895–1915
42. Hung M-H, Hsieh C-H, Kuo C-M (2006) Similarity retrieval of shape images based on database classification. J Vis Commun Image Represent 17:970–985
43. Jain AK, Vailaya A (1998) Shape-based retrieval: a case study with trademark image databases. Pattern Recognit 31(9):1369–1390
44. Joly A, Buisson O (2009) Logo retrieval with a contrario visual query expansion. In: Proceedings of the 17th ACM international conference on multimedia (MM '09), Beijing, pp 581–584
45. Kim YS, Kim WY (1998) Content-based trademark retrieval system using a visually salient feature. Image Vis Comput 16:931–939
46. Kleban J, Xie X, Ma W-Y (2008) Spatial pyramid mining for logo detection in natural scenes. In: Proceedings of the IEEE conference on multimedia expo, Hannover, pp 1077–1080
47. Levenshtein V (1966) Binary codes capable of correcting deletions, insertions, and reversals. Cybern Control Theory 10(8):707–710
48. Li Z, Schulte-Austum M, Neschen M (2010) Fast logo detection and recognition in document images. In: Proceedings of the 20th international conference on pattern recognition, Istanbul, pp 2716–2719
49. Logo Dataset (2012) Laboratory for Language and Media Processing (LAMP), University of Maryland. http://lamp.cfar.umd.edu
50. Lowe D (2004) Distinctive image features from scale-invariant keypoints. IJCV 60(2): 91–110
51. Lowe DG (1985) Perceptual organization and visual recognition. Kluwer Academic, Boston
52. Luo J, Crandall D (2006) Color object detection using spatial-color joint probability functions. IEEE Trans Image Process 15(6):1443–1453
53. Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, New York
54. Meisinger K, Troeger T, Zeller M, Kaup A (2005) Automatic TV logo removal using statistical based logo detection and frequency selective inpainting. In: Proc. European signal processing conference. Antalya, Turkey, 4–8
55. Meng J, Yuan J, Jiang Y, Narasimhan N, Vasudevan V, Wu Y (2010) Interactive visual object search through mutual information maximization. In: Proceedings of the ACM international conference on multimedia, Firenze, pp 1147–1150
56. Mori G, Belongie S, Malik J (2005) Efficient shape matching using shape contexts. IEEE Trans Pattern Anal Mach Intell 27(11):1832–1837
57. Neumann J, Samet H, Soffer A (2002) Integration of local and global shape analysis for logo classification. Pattern Recognit Lett 23(12):1449–1457
58. Oliva A, Torralba A (2007) The role of context in object recognition. Trends Cogn Sci 11:520–527
59. Ozay N, Sankur B (2009) Automatic TV logo detection and classification in broadcast videos. In: EUSIPCO 2009, Glasgow, pp 839–843
60. Phan R, Androutsos D (2009) Content-Based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms. Comput Vis Image Underst 114(1):66–84
61. Pham T (2003) Unconstrained logo detection in document images. Pattern Recognit 36(12):3023–3025
62. Quack T, Ferrari V, Liebe B, Gool LV (2007) Efficient mining of frequent and distinctive feature configurations. In: ICCV, Rio de Janeiro
63. Ren M, Eakins JP, Briggs P (2000) Human perception of trademark images: implications for retrieval system design. J Electron Imaging 9(4):564–575

64. Rocchio JJ Jr (1971) Relevance feedback in information retrieval. The smart system-experiments in automatic document processing. Prentice-Hall, Englewood Cliff, pp 313–323
65. Rusinol M, Llados J (2009) Logo spotting by a bag-of words approach for document categorization. In: ICDAR'09, Barcelona, pp 111–115
66. Rusiñol M, Lladós J (2010) Efficient logo retrieval through hashing shape context descriptors. In: Proceedings of the 9th international workshop on document analysis systems, Boston, pp 215–222
67. Rusiñol M, Nourbakhsh F, Karatzas D, Valveny E, Lladós J (2010) Perceptual image retrieval by adding color information to the shape context descriptor. In: Proceedings of the 20th international conference on pattern recognition, Istanbul. IEEE, pp 1594–1597
68. Rusiñol M, Aldavert D, Karatzas D, Toledo R, Lladós J (2011) Interactive trademark image retrieval by fusing semantic and visual content. In: Advances in information retrieval: 33rd European conference on IR research, Dublin. Lecture notes in computer science, vol 6611, pp 314–325
69. Santos AR, Kim HY (2006) Real-Time opaque and semi-transparent TV logos detection. In: Proceedings of the 5th international information and telecommunication technologies symposium (I2TS), Cuiabá
70. Sarkar S, Boyer KL (1994) Computing perceptual organization in computer vision. Series in machine perception and artificial intelligence, vol 12. World Scientific, Singapore/River Edge
71. Saund E (2003) Finding perceptually closed paths in sketches and drawings. IEEE Trans Pattern Anal Mach Intell 25(4):475–491
72. Saund E (2011) PPD: platform for perceptual document analysis. PARC TR-2011-1, Nov 2011
73. Schietse J, Eakins JP, Veltkamp RC (2007) Practice and challenges in trademark image retrieval. In: Proceedings of the 6th ACM international conference on image and video retrieval, Amsterdam, pp 518–524
74. Seiden S, Dillencourt M, Irani S, Borrey R, Murphy T (1997) Logo detection in document images. In: Proceedings of the international conference on imaging science, systems, and technology, Las Vegas, Nevada, USA, pp 446–449
75. Sivic J, Zisserman A (2003) Video Google: a text retrieval approach to object matching in videos. In: ICCV, Nice, pp 1470–1477
76. The legacy tobacco document library (LTDL) at UCSF (2006). http://legacy.library.ucsf.edu
77. Tombre K, Lamiroy B (2003) Graphics recognition – from re-engineering to retrieval. In: Proceedings of the seventh international conference on document analysis and recognition, ICDAR03, Edinburgh, pp 148–155
78. Tschumperle D (2006) Fast anisotropic smoothing of multi-valued images using curvature-preserving PDE's. Int J Comput Vis 68(1):65–82
79. van de Sande KEA, Gevers T, Snoek CGM (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell 32(9):1582–1596
80. Venters CC, Hartley RJ, Cooper MD, Hewitt WT (2001) Query by visual example: assessing the usability of content-based image retrieval system user interfaces. In: Shum H-Y, Liao M, Chang S-F (eds) PCM 2001, Beijing. LNCS 2195, pp 514–521
81. Wang H, Chen Y (2009) Logo detection in document images based on boundary extension of feature rectangles. In ICDAR'09, Barcelona, pp 1335–1339
82. Wang J, Duan L, Li Z, Liu J, Lu H, Jin JS (2006) A robust method for TV logo tracking in video streams. In: Proceedings of the IEEE international conference on multimedia and expo (ICME), Toronto, pp 1041–1044
83. Watve A, Sural S (2008) Soccer video processing for the detection of advertisement billboards. Pattern Recognit Lett (29):994–1006
84. Wei C-H, Li Y, Chau W-Y, Li C-T (2009) Trademark image retrieval using synthetic features for describing global shape and interior structure. Pattern Recognit 42:386–394
85. Wertheimer M (1938) Untersuchungen zur Lehre der Gestalt, II. Psychologische Forschung, vol 4, pp 301–350, 1923. Translation published as Laws of organization in perceptual forms. In: Ellis W (ed) A source book of Gestalt psychology, Routledge and Kegan Paul, London, pp 71–88

86. Witkin AP, Tenenbaum JM (1982) On the role of structure in human and machine vision. In: Beck J, Hope B, Rosenfeld A (eds) Human and machine vision. Academic Press, New York, pp 481–543
87. World Intellectual Property Organization (2007) International classification of the figurative elements of Marks (Vienna classification), 6th edn. WIPO publication No. 502E/6, Geneva, Academic Press, New York
88. Wu JK, Lam CP, Mehtre BM, Gao YJ, Desai Narasimhalu A (1996) Content based retrieval for trademark registration. Multimed Tools Appl 3(3):245–267
89. Yan WQ, Wang J, Kankanhalli MS (2005) Automatic video logo detection and removal. Multimed Syst 10(5):379–391
90. Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recognit 37:1–19
91. Zhu G, Doermann D (2007) Automatic document logo detection. In: Proceedings of the international conference on document analysis and recognition, Curitiba, pp 864–868

Further Reading

A number of key references for each of the applications discussed in this chapter are summarized in Tables 18.3–18.6. The cited work is selected to cover all the various aspects of the discussed applications; hence, the proposed lists comprise a good starting point for readers interested in a broad review of the state of the art.

Readers with a specific interest in perceptual organization are encouraged to consult [51] and [21] while the "Platform for Perceptual Document Analysis" would provide further practical insight [72]. For a comprehensive overview of information retrieval systems, the interested researcher can consult [7] and [53], while for a more practical analysis, [36] offers a good starting point. Finally, a recent work dedicated specifically to logo recognition [16] provides an introduction to fundamental concepts and methods in pattern and shape recognition while surveying advances in the area.