Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 39 (2006) 988–1001
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Gestalt-based feature similarity measure in trademark database

Hui Jiang[a, b], Chong-Wah Ngo[a,*], Hung-Khoon Tan[a]

[a]*Department of Computer Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*
[b]*Institute for Computational and Mathematical Engineering, Stanford University, USA*

Received 29 September 2004; received in revised form 5 July 2005; accepted 15 August 2005

Abstract

Motivated by the studies in Gestalt principle, this paper describes a novel approach on the adaptive selection of visual features for trademark retrieval. We consider five kinds of visual saliencies: symmetry, continuity, proximity, parallelism and closure property. The first saliency is based on Zernike moments, while the others are modeled by geometric elements extracted illusively as a whole from a trademark. Given a query trademark, we adaptively determine the features appropriate for retrieval by investigating its visual saliencies. We show that in most cases, either geometric or symmetric features can give us good enough accuracy. To measure the similarity of geometric elements, we propose a maximum weighted bipartite graph (WBG) matching algorithm under transformation sets which is found to be both effective and efficient for retrieval.

Keywords: Trademark image retrieval; Gestalt principle; Bipartite graph matching under transformation sets

1. Introduction

To date, despite the numerous efforts in content-based image retrieval (CBIR), finding the best shape features and the best way of matching features for image retrieval remains challenging. One of the core issues is in formulating a general-purpose shape similarity measurement that guarantees good retrieval performance, with the baseline that the retrieved similar items should be consistent with human visual perception. Recently, Gestalt principle [1] is taken into account by researchers for the perceptual segmentation and grouping of shape features. Gestalt principle is one of the earliest studies conducted by a group of psychologists to model shape perception in the early 19th century. A number of principles have been experimentally studied and derived to govern the grouping of shape features.

Perception, in general, is viewed as an active process of organization, construction and analysis. Gestalt principle emphasizes the *holistic* nature, where recognition is inferred more by the properties of an image as a whole, rather than parts, during visual perception. This is considered different from traditional pattern recognition where recognition is achieved by accounting image features of parts and their combinations. Take the image in Fig. 3 as an example. Gestalt principle considers white regions (areas enclosed by five group of parallel lines) as a whole as the significant property rather than the shape of six independent black regions.

In this paper, we investigate the flexibility of applying Gestalt principle in trademark database since trademarks are images that usually contain rich *abstract* geometric features that are appropriate for the modeling of Gestalt principle. In particular, we focus on five *holistic* properties: symmetry, continuity, proximity, parallelism and closure derived in Gestalt principles. The first property is described by Zernike moments, while the others are extracted and represented *illusively*[1] as a whole by our proposed geometrical features under the weighted bipartite graph (WBG) framework [2–6]. These five holistic properties, in general, are not effective

* Corresponding author. Tel.: +852 2784 4390; fax: +852 2788 8614.
E-mail address: cwngo@cs.cityu.edu.hk (C.-W. Ngo).

[1] We use the word "illusively" to describe the nature of Gestalt principles and the motivation of our approach: Human always group low-level geometrical elements illusively as one or several "complete elements", even though a complete element is actually not connected and formed by several broken segments. For example, in Fig. 5b, the inner circle is broken into three arcs, but our approach can detect them as a "whole" (a complete circle) which mimics the human perceptual organization.

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.08.012

if they are jointly integrated in a linear weighted combination way for retrieval. To solve this problem, we propose a novel adaptive selection procedure of holistic properties, which depends on the nature of a query image.

Gestalt principle has been investigated in [7–13] for trademark retrieval. However, only a subset of holistic properties is utilized. No study has yet been carried out on how to systematically select and match these properties for trademark retrieval. In [7–13], clustering algorithms are employed to group semantically meaningful Gestalt components. After clustering, non-geometric features such as aspect ratio, circularity and degree of perpendicularity are extracted from each cluster for retrieval. Nevertheless, as pinpointed in [13], the incorrect clustering of elements is the major drawback that affects the retrieval accuracy. In this paper, instead of adopting clustering based approach, we encode directly the extracted geometric elements led by Gestalt principle in WBG for partial matching under a set of allowable transformations. Since our approach matches geometric elements as a whole directly, it leads to a more reliable framework for trademark similarity measurement.

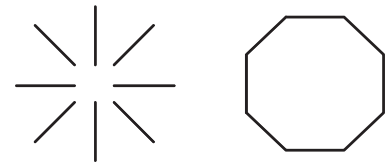

Fig. 1. An example showing the weakness of the non-geometric features: these two trademarks have similar edge direction histograms, but they are quite different.

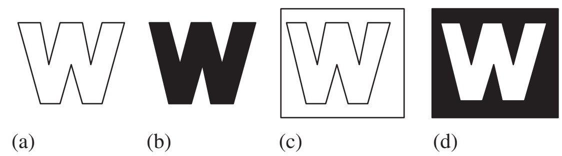

Fig. 2. An example showing the weakness of moments.

2. Related works

Numerous approaches have been proposed for trademark image retrieval. Representative works include [7–17]. As other CBIR problems, most approaches in trademark image retrieval consist of two major components: feature extraction and similarity measurement. The choice of features will normally affect the use of similarity measurement. In this section, we focus our attention on how they derive the shape features for similarity retrieval.

In the current literature, various visual features have been explored for trademark retrieval. The features adopted most frequently are: edge direction histogram [16,18], moments [16–18] and shape descriptors [7–12,19,20]. Some of these features (e.g., edge direction histogram and moments) contain no geometric information. They are global and statistical in some sense. For convenience, we call them non-geometric features. Since no geometrical information is encoded in these features, two images with similar features can be very different (See Fig. 1 for example). Similarly, two images with similar shape may have considerably different global and statistical features. One example is given in Fig. 2, the four trademarks are perceptually similar to each other but their moments are very different. In [16], Jain and Vailaya proposed an image filling approach to solve this problem. However, this method can only handle the cases of Fig. 2(a) and (b). Because the moments do not fit human perception very well, recent approaches in [7–12,20] consider only the edge points of regions for trademark retrieval.

Although the non-geometric features have the weakness introduced above, they have the advantages of being easy to compute and compare. Most importantly, the weakness may be attenuated by integrating multiple features (a straightforward observation from Figs. 1 and 2). Nevertheless, the experimental results conducted by Eakin et al. in [12] indicated that it is not always true that the combination of multiple features can give better results than using them on their own. The key issue is how to effectively integrate multiple features, which is not a trivial problem. In [16], Jain and Vailaya employed a two-level hierarchical system. In the first stage, edge direction histograms and moments were used to rapidly filter the database. In the second stage, deformable template matching was used for final similarity ranking. The reasons why they used such a framework are: edge direction histograms and moments are non-geometrical features, they are quick but coarse; deformable template matching takes into account the geometric information, but it is accurate but slow. The experimental results in [16] showed that moments are not robust to trademarks with line drawings, and the deformable matching is not effective for the trademarks with many details in line drawings and holes. Their results are improved by filling-in the holes in the trademarks, but the major drawback is the non-utilization of information in holes. For example, the trademark in Fig. 2(d) becomes a square after image filling, the shape information "W" is missed after filling.

Instead of extracting the global features as a whole from the images as in [16,17], there is a more general scheme in [7–12,19,20]: decompose the images into several components, and then use non-geometric features to encode each component. Decomposition of trademark images is a hard problem. In [19–22], trademarks are segmented into regions based on the pixel connectivity and the shape features are extracted from each region for retrieval. The segmentation by pixel connectivity, nevertheless, does not always reflect the segmentation by human. Consider the trademark shown in Fig. 3, the shape of this trademark is inferred as a whole from the image, rather than from each individual region. To segment a trademark into perceptually meaningful components, Gestalt principle [1] is taken into account in [7–13].