Available online at www.sciencedirect.com

ScienceDirect

Journal of Visual Languages and Computing 19 (2008) 637–651

Journal of Visual Languages & Computing

www.elsevier.com/locate/jvlc

# Spatial inference and similarity retrieval of an intelligent image database system based on object's spanning representation

Po-Whei Huang[a,*], Lipin Hsu[b], Yan-Wei Su[a], Phen-Lan Lin[c]

[a]*Department of Computer Science, National Chung Hsing University, Taichung, Taiwan*
[b]*Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan*
[c]*Department of Computer Science and Information Management, Providence University, Salu, Taiwan*

Received 12 October 2006; received in revised form 7 August 2007; accepted 25 September 2007

**Abstract**

In this paper, we presented a novel image representation method to capture the information about spatial relationships between objects in a picture. Our method is more powerful than all other previous methods in terms of accuracy, flexibility, and capability of discriminating pictures. In addition, our method also provides different degrees of granularity for reasoning about directional relations in both 8- and 16-direction reference frames. In similarity retrieval, our system provides twelve types of similarity measures to support flexible matching between the query picture and the database pictures. By exercising a database containing 3600 pictures, we successfully demonstrated the effectiveness of our image retrieval system. Experiment result showed that 97.8% precision rate can be achieved while maintaining 62.5% recall rate; and 97.9% recall rate can be achieved while maintaining 51.7% precision rate. On an average, 86.1% precision rate and 81.2% recall rate can be achieved simultaneously if the threshold is set to 0.5 or 0.6. This performance is considered to be very good as an information retrieval system.


*Keywords:* Image database; Image representation; Spatial inference; Similarity retrieval

## 1. Introduction

Pictorial databases have been used more widely in recent years. In applications such as computer-aided design, office automation, medical image archiving, geographic information, and trademark picture registration, retrieving pictures from pictorial databases is frequently required. Content-based image retrieval is the current trend of designing image database systems as opposed to text-based image retrieval [1–9]. Rather than proceeding via a manually generated text-based description, the content-based image retrieval works by matching the query image against a database image according to the contents of images.

The features used in content-based image retrieval can be roughly divided into two categories: the low-level visual features such as color, shape, and texture and the high-level features such as spatial relationships among the objects in a picture. Examples of content-based image retrieval systems include QBIC [10], VisualSEEK [11], and WaveGuide [12], etc. They allow users to retrieve similar

*Corresponding author. Tel.: +886 4 2235 0813.
*E-mail address:* pwhuang@cs.nchu.edu.tw (P.-W. Huang).

1045-926X/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.jvlc.2007.09.001

pictures from a large database based on low-level visual features.

Retrieving pictures that satisfy high-level spatial queries is also an important issue in image database systems [13–20]. For example, to answer the query "find all pictures having a swimming pool to the left of a house", we need to keep at least the directional relationship between the swimming pool and the house for all pictures.

In this paper, we present a spatial knowledge representation method for pictures containing nonzero-sized objects. The spatial relations we considered are both topological and directional relations between nonzero-sized objects. A set of similarity measures for handling different types of queries that will satisfy various kinds of users' requirements is also proposed. We use some examples to demonstrate that our image representation method is more powerful than many well-known methods. In particular, we built a database of 3600 pictures, with each picture containing four to six objects, to demonstrate the performance of our system in terms of recall and precision. Experiment results show that, in the two extreme cases, 97.8% precision rate can be achieved with 62.5% recall rate, and 97.9% recall rate can be achieved with 51.7% precision rate. On an average, a very good performance of 86.1% precision rate and 81.2% recall rate can be achieved simultaneously with threshold $T_h = 0.5$ or 0.6.

## 2. Overview

Spatial relationships between objects have been identified as one of the most important features for describing the contents of images. Unlike topological relations which are well defined and less disputable, directional relations can be viewed and modeled in different ways. Previous methods for modeling directional relations include direction between Minimum Bounding Rectangles (MBRs) [13,18–20], 2D-strings [16,17], 2D-PIR [22], triangular model [23], and symbolic arrays [24]. A detailed analysis for comparing these methods can be found in Ref. [24]. Except the above methods, two other direction models, direction–relation matrix [21] and 9D-SPA [14], based on rectangle-shaped partition were also proposed. They partition the whole space around a reference object and record into which direction tiles an target object falls. They can provide better approximations for spatial relations between objects with complex structures including shapes such as concave regions or objects with holes. The 9D-SPA method can even support an efficient indexing structure to facilitate search in similarity retrieval. In this paper, we only concentrate on iconic picture retrieval based on spatial relations where the set of icons are known and each object in a picture must match an icon [25].

Different from 9D-SPA which uses rectangle-shaped partition, this paper presents an image representation method for capturing the relations among nonzero-sized objects based on a triangular partition approach. Triangular partition is more consistent with human perception and is better than rectangle-shaped partition from an observer's point of view, if the observer is placed at the position where the reference object is located. Such a phenomenon is clearly demonstrated in Fig. 1, where object $A$ is to the east of object $B$ in triangular partition while object $A$ is to the northeast of object $B$ in rectangle-shaped partition. However, directions based on triangular partition approach may still be misleading and not suitable for direction queries in spatial databases if objects are overlapping, intertwined, or horseshoe-shaped. This paper improves

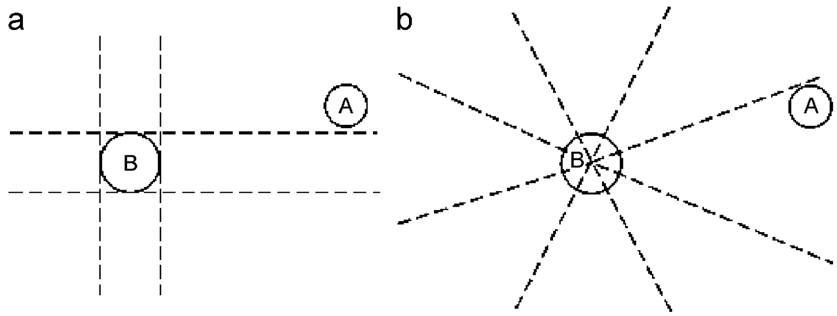


Fig. 1. $A$ is the target object and $B$ is the reference object. (a) Rectangle-shaped partition: $A$ is to the northeast of $B$. (b) Classic triangular partition: $A$ is to the east of $B$.