The Journal of Systems and Software 82 (2009) 685–696

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

A new relevance feedback technique for iconic image retrieval based on spatial relationships

Peng-Yeng Yin *, Chin-Wen Liu

Department of Information Management, National Chi-Nan University, Nantou 54561, Taiwan

ARTICLE INFO

Article history:
Received 21 May 2008
Received in revised form 9 September 2008
Accepted 30 September 2008
Available online 11 October 2008

Keywords:
2D String
Longest common subsequence
Shortest common supersequence
Relevance feedback

ABSTRACT

Due to the popularity of Internet and the growing demand of image access, the volume of image databases is exploding. Hence, we need a more efficient and effective image searching technology. Relevance feedback technique has been popularly used with content-based image retrieval (CBIR) to improve the precision performance, however, it has never been used with the retrieval systems based on spatial relationships. Hence, we propose a new relevance feedback framework to deal with spatial relationships represented by a specific data structure, called the 2D B_e-string. The notions of relevance estimation and query reformulation are embodied in our method to exploit the relevance knowledge. The irrelevance information is collected in an irrelevant set to rule out undesired pictures and to expedite the convergence speed of relevance feedback. Our system not only handles picture-based relevance feedback, but also deals with region-based feedback mechanism, such that the efficacy and effectiveness of our retrieval system are both satisfactory.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Traditional image retrievals rely on *passive query* that searches the desired targets by submitting alphanumerical terms to be matched to a set of filenames or keywords describing the stored items. This was considered a crippled process that invites at least three drawbacks. (1) It is laborious to annotate manually every stored image by a name or a list of keywords because modern databases easily contain thousands of items. (2) Many images involve multiple visual and semantic concepts, it is limited to annotate such images by a small number (usually less than five) of text terms. (3) Inconsistencies and ambiguities exist between the text terms that different people use to describe a given image. As a departure from these limitations, content-based image retrieval (CBIR) (Lew et al., 2006; Datta et al., 2008) advocates *active query* that retrieves relevant images by providing a piece of visual information which has been implemented in various ways by several systems (e.g., Photobook (Pentland et al., 1996), Surfimage (Nastar et al., 1998), SQUID (Mokhtarian et al., 1996), MPE (Vasconcelos, 2004), PicSOM (Iivarinen et al., 2004), just to name a few) using image samples, object sketches, or other visual features (color, shape, texture, spatial relationships, etc.) CBIR systems are easy to use, however, the retrieval precision may be unsatisfactory due to imperfect features generated in the image processing stage and the gap between the visual and semantic concepts.

A remedy to this problem is to treat the retrieval session as repetitive query reformulation operations. Through successive human-computer interactions, the query descriptive information (features, matching models, metrics or any meta-knowledge) is re-modified as a response to the user's feedback on retrieved results. Therefore, an approximate "ideal" query and an adaptively tuned matching model are eventually produced and the system is able to attain high retrieval precision. Relevance feedback (RF) is a query reformulation process that was originally employed in information retrieval applications (Rocchio, 1971) for accuracy improvement and RF has been applied to CBIR systems since late 1990's (Ciocca and Schettini, 1999; Rui et al., 1998; Peng et al., 1999; Meilhac and Nastar, 1999; Cox et al., 2000; Su et al., 2003; Tong and Chang, 2001; Tieu and Viola, 2004; Li and Hsu, 2008; Hoi and Lyu, 2004; Yin et al., 2008, 2005). The basic process of RF under the CBIR scenario proceeds as follows. The user initializes a query session by submitting a relevant image descriptor (e.g., an image sample, a draft sketch, or visual feature tiles). The system then uses this descriptor to search a list of top v relevant images in the database. If the user is not satisfied with the retrieved result, he/she can call for another retrieving iteration by specifying which currently returned images are relevant and which are irrelevant. The system then modifies the query descriptor and system parameters appropriately, so more user-oriented images will be included in the next retrieved result. This process is repeated until the user is satisfied or the results cannot be further improved.

Although the RF techniques have been applied to several CBIR applications, it's usage has not been exploited yet for improving

* Corresponding author. Tel.: +886 49 2910960; fax: +886 49 2915205.
E-mail address: pyyin@ncnu.edu.tw (P.-Y. Yin).

0164-1212/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2008.09.033

the retrieval performance of systems based on spatial information. One major reason is that the visual representation used by the latter is significantly different from those based on other visual features. Most CBIR systems use numerical visual features such as color, shape, or texture, this representation well fits the traditional RF vector model employed in information retrieval, making them easy to marry and to create benefits. However, the prevailing index scheme for spatial information retrieval is the symbolic 2D string (Chang et al., 1987) and it's extensions (Chang and Li, 1988; Jungert and Chang, 1989; Lee and Hsu, 1990; Lee et al., 1992; Petraglia et al., 1993; Wang, 2003; Lee and Chiu, 2003). There is a wealth of challenging issues deserve studying to innovate new RF techniques that can handle symbolic strings for spatial information retrieval. To the best of our knowledge, this paper is the first attempt that aims to propose a new RF framework that is able to derive implicit spatial relevance by exploiting user's historical feedback.

The remainder of this paper is organized as follows. Section 2 reviews the spatial information retrieval and traditional RF techniques. Section 3 describes the proposed RF framework for retrievals based on spatial similarity. Section 4 presents the experimental results and performance evaluation. Finally, conclusions are given in Section 5.

2. Related work

We review the representation schemes, namely, the 2D string and it's extensions, for spatial information and the similarity matching between iconic images represented by such schemes. Then, the traditional RF models for CBIR applications using numerical visual features are summarized.

2.1. Retrieval by spatial similarity

2.1.1. Representation of spatial information

Spatial information is an important clue for determining relevant pictures in many applications such as geographic information system, urban planning, indoor decoration design, etc. Chang et al. (1987) originally proposed a novel data structure, called 2D string, for representing the spatial information among objects contained in an iconic picture. The 2D string tallies the sequence of object projections along the x and y axes, respectively. Several spatial relation indicators are introduced to describe the positional relationships. Another representation method for describing the spatial information is the 9DLT (Directional Lower Triangular) scheme (Chang and Jiang, 1994). The 9DLT scheme uses a triplet (O_i, O_j, d_{ij}) to tally the spatial relationship between the reference object O_i and a target object O_j by a directional code d_{ij}. All the triplets generated from the iconic pictures are transformed into a signature file on which the future retrieval is performing. The 9DLT scheme provides quick access to target pictures by reference to the signature file. However, it cannot support similarity ranking on the retrieved result, which is the core concept of the relevance feedback. As a consequence, we did not consider the 9DLT scheme in this paper.

Fig. 1 shows an example of the 2D string for an iconic picture whose objects have been represented by alphabetical symbols put in sets separated by columns and rows. The symbolic projection along the x-axis is described by the sequence (A = D:E < B < C) while the projection along the y-axis is indicated by (A < B = C < D:E). The spatial relation indicator "=" means "at the same location as", the indicator "<" denotes "left of" and "below", and the indicator ":" corresponds to the spatial relationship "in the same set as". Thus, a 2D string comprising the two projection sequences renders the spatial information involved in an iconic picture.

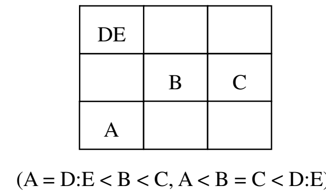

Fig. 1. The 2D string of an iconic picture.

The 2D string, nevertheless, is not able to describe complex spatial relationships such as "containing" and "partial overlapping". It invites many researchers to develop variants for enhancing the descriptive capability of the 2D string family. Instead of reviewing each of the 2D string variants in detail, the major features of them are summarized in Table 1. Here, we review an instance, namely, the 2D B_e-string (Wang, 2003), because it is particularly suited to relevance matching using spatial information and it is thus adopted in our proposed RF approach.

Wang (2003) proposed the 2D B_e-String. In this representation scheme, all of the spatial relation indicators are removed, hence providing a simpler and more efficient form. On each axis, every object produces two projection lines intersecting at the beginning and ending boundaries of the object along the corresponding axis. The 2D B_e-String is produced by tallying the sequence of the object projection lines from left to right and from bottom to top for the x and y axes, respectively. A dummy object projection line "e" is inserted between two consecutive projection lines if the two projection lines are not coinciding, so the ambiguity of positional coincidence incurred in spatial matching can be resolved.

Fig. 2 illustrates an instance of an iconic picture containing four objects labeled as A, B, C and D and the corresponding 2D B_e-string is shown at the bottom of the picture. The subscripts "b" and "e" of object labels denote the beginning and ending projection lines, respectively. Consequently, the 2D B_e-string of the picture is read as ($eA_bB_beA_eeB_eeD_beC_beC_eeD_ee$, $eB_bD_beB_eeD_eeA_beC_beA_eeC_ee$).

We adopted the 2D B_e-string as our representation scheme because it is suited to similarity matching based on common subsequence which supports similarity ranking on the retrieved result. The other 2D string variants using spatial indicators would cause ambiguity in subsequence matching. For example, the longest common subsequence between "A = B < C" and "A = C < D" is "A = C", however, the actual spatial relationship between object A and C in the first string is "A < C", which contradicts with the result of the subsequence matching. The 2D B_e-string scheme inserts a dummy character "e" between two consecutive projection lines if the two projection lines are not coinciding, so the ambiguity incurred in the subsequence matching can be resolved.

2.1.2. Similarity matching between iconic pictures

The similarity retrieval using spatial information can be implemented by defining a distance metric between the spatial-relation strings whichever are used to represent the iconic pictures. To this end, there are two main types of distance metrics proposed in the literature, namely, the rank-based distance metric (Chang et al., 1987; Chang and Lee, 1991; Lee and Hsu, 1992) and the longest common subsequence (LCS)-based distance metric (Wang, 2003; Lee et al., 1989). Chang et al. (1987) defined the rank-based distance metric which identifies type-i ($i = 0, 1, 2$) 2D subpictures from the most relaxed matching to the most stringent matching based on the positional ranks of objects. However, the matching based on type-i subpictures requires that all the objects appear in the query picture must be also present in the retrieved database pictures. This mechanism forbids the possibility for facilitating partial matching which allows the retrievals of database pictures that contain some but not all of the objects in the query picture. To this

Download English Version:

https://daneshyari.com/en/article/462250

Download Persian Version:

https://daneshyari.com/article/462250

Daneshyari.com