Up-to-Date Summary of Semantic-Based Visual Information Retrieval

Yu-Jin Zhang
Department of Electronic Engineering, Tsinghua University, China

INTRODUCTION

Content-based image retrieval (CBIR) has made its first appearance about 20 years ago. It could be described as a process framework for efficiently retrieving images from a collection by similarity. The retrieval relies on extracting the appropriate characteristic quantities describing the desired contents of images. Content-based video retrieval (CBVR) made its appearance in treating video in the similar means as CBIR treating images. Content-based visual information retrieval (CBVIR) combines CBIR and CBVR together (Zhang, 2003).

Along with the progress of electronic equipments and computer techniques for visual information capturing and processing, a huge number of image and video records have been collected. Visual information becomes a well-known information format and a popular element in all aspects of our society. The large visual data make the dynamic research to be focused on the problem of how to efficiently capture, store, access, process, represent, describe, query, search, and retrieve their contents. In the last years, CBVIR has experienced significant growth and progress, resulting in a virtual explosion of published information. It has attracted many interests from image engineering, computer vision and database community.

In recent years, the focus of CBVIR is around capturing high-level semantics, i.e., the so-called Semantic-Based Visual Information Retrieval (SBVIR). This article will first show some statistics about the research publications on SBVIR in recent years to give an idea about its developments statue. It then gives an overview on several current centers of attention, by summarizing results on subjects such as image and video annotation, human-computer interaction, models and tools for semantic retrieval, and miscellaneous techniques in applications. Finally, some future research directions, the domain knowledge and learning, relevance feedback and association feedback, as well as research at even high levels, such as cognitive level and affective level are pointed out.

BACKGROUND

To get a general idea about the scale and progress of research on CBVIR and SBVIR for the past years, several searches in EI Compendex database (http://www.ei.org) for papers published in English from 1993 through 2012 (totally 20 years) have been made. In Table 1, the results of two searches in the title field for the numbers of English published papers (records) are listed, one term used is “image retrieval (IR)” and other term is “semantic image retrieval (SIR).” The papers found out by the second term should be a sub-set of the papers found out by the first term. Both numbers have a tendency of increasing in general for these 20 years, with some stabilization in the last five years, as can be seen clearly from Table 1.

Other searches take the same terms as used for Table 1, but are performed in the field of title/abstract/subject. The results are shown in Table 2. Both numbers have a similar tendency of increasing, especially in the first 15 years, and then a general slowly-decrease can be observed.

One analysis to the Table 1 and Table 2 is comparing the ratios of SIR over IR in two tables. It is seen that these ratios in Table 2 are much higher than those ratios in Table 1. This difference indicates that the research for SIR is still in an early stage (many papers have not put word “semantic” in the title of papers) but this concept starts to get numerous considerations or attracts many attentions (“semantic” appeared already in abstract and/or subject parts of these papers).

To have a more close comparison, these ratios in Table 1 (called Ratio A) and Table 2 (called Ratio B)

DOI: 10.4018/978-1-4666-5888-2.ch123

Table 1. List of English records found in the title field of EI Compendex for the last 20 years

Searching Terms	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	
(1) Image Retrieval	25	20	70	89	80	131	155	161	191	233	
(2) Semantic Image Retrieval	0	0	0	1	1	2	4	5	4	8	
Ratio of (2) over (1)	0	0	0	1.12	1.25	1.53	2.58	3.11	2.09	3.43	
Searching Terms	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**Total**
(1) Image Retrieval	241	358	417	418	476	605	586	615	535	612	6018
(2) Semantic Image Retrieval	11	18	30	28	37	44	36	37	36	40	342
Ratio of (2) over (1)	4.56	5.03	7.19	6.70	7.77	7.27	6.14	6.02	6.73	6.54	5.68

Table 2. List of English records found in the subject/title/abstract field of EI Compendex for the last 20 years

Searching Terms	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	
(1) Image retrieval	273	288	421	580	531	640	718	871	1080	1203	
(2) Semantic image retrieval	0	9	11	25	27	45	62	79	109	131	
Ratio of (2) over (1)	0	0.03	2.61	4.31	5.08	7.03	8.64	9.07	10.09	10.89	
Searching Terms	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**Total**
(1) Image retrieval	1267	2196	2174	2697	3025	4541	4842	3414	3393	3375	37529
(2) Semantic image retrieval	153	324	257	274	378	485	443	434	385	379	4010
Ratio of (2) over (1)	12.08	14.75	11.82	10.16	12.50	10.68	9.15	12.71	11.35	11.22	10.69

are plotted together in Figure 1. In Figure 1, dark bars represent ratios from Table 1 and light bars represent ratios from Table 2. In addition, the tendencies of ratio developments are approximated by third order polynomials. It is clear that many papers have the "semantic" concept in mind though they do not always use word "semantic" in the title. In addition, with the grown-up of research for semantic image retrieval, these two ratios have the trend of close in recent years.

MULTI-LEVEL SEMANTIC IMAGE RETRIEVAL

The motivation for the research on semantic image retrieval comes from the problem of "semantic gap." In the early stage of content-based image retrieval, contents are represented by various visual features (for example, color features, texture features, shape features, space relationship features, etc.). The problem of the feature-based technique is that there is a considerable difference between the users' interesting in reality and the image contents described by only using the above low-level perceptive features (Zhang et al., 2004), though all current techniques assume certain mutual relations and cross-information between the similarity measures and the semantics of images and videos. In other words, there is a large gap between such content description based on low-level features and that of human beings' understanding. As a result, these feature-based approaches often lead to unsatisfying querying results in many different cases (Smeulders et al., 2000).

To fill the "semantic gap," one solution is to make the retrieval system working with low-level features while the user puts in high-level knowledge, so as to map low-level visual features to high-level semantics (Zhou & Huang, 2002). Two typical early methods are to optimize query request by using relevance feedback and semantic visual template (Chang et al., 1998) and to interpret progressively the content of images by using interactive interface (Castelli et al. 1998).

Nowadays, the mainstream of the research converges to retrieval based on semantic meaning, which tries to extract cognitive concept of human by combining

Related Content