Using Context Information to Improve Retrieval Accuracy in Content-Based Image Retrieval Systems

Mahmoud R. Hejazi, Woontack Woo, Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
Oryong Dong, Buk gu, Gwangju 500-712, South Korea
{m_hejazi, wwoo, hoyo}@gist.ac.kr

Abstract

Current image retrieval techniques have shortcomings that make it difficult to search for images based on a semantic understanding of what the image is about. Since an image is normally associated with multiple contexts (e.g. when and where a picture was taken,) the knowledge of these contexts can enhance the quantity of semantic understanding of an image. In this paper, we present a context-aware image retrieval system, which uses the context information to infer a kind of metadata for the captured images as well as images in different collections and databases. Experimental results show that using these kinds of information can not only significantly increase the retrieval accuracy in conventional content-based image retrieval systems but decrease the problems arise by manual annotation in text-based image retrieval systems as well.

Keyword : Image Retrieval, Context-Aware Computation, Annotation

1. Introduction

In recent years, image collections have increased both in numbers and in size, and represent a huge amount of important information. Both privately and publicly owned collections of images are available over the Internet for browsing and searching, and the number of users and application areas are increasing. To access these image collections efficiently based on the users' needs, which are usually available in an abstract notion, powerful content-based image retrieval (CBIR) techniques are crucial [1].

However, the retrieval accuracy of conventional CBIR systems is often low since the retrieval is usually performed based on the comparison of low level features such as color, shape, and texture. Furthermore, many practical CBIR systems still rely on text retrieval technologies on human labeled keywords which are almost impractical when we deal with increasingly large amount of image collections.

We can expect to increase the retrieval accuracy if we use semantics (metadata) for expressing the images. One well-known standard in this respect is MPEG-7 that provides a framework for definition of metadata to describe the content of media objects, principally audio-visual objects [2]. Although the idea of using semantics for images is an interesting and challenging topic in the field of image retrieval, there are still many problems for proper understanding and annotating such semantics automatically [3] [4].

In this paper, we present a convenient and efficient method to handle this problem, considering the fact that nowadays we are easily able to sense, infer, and learn the context of creation and use of a media (e.g. an image) using very advanced capturing devices. In this respect, we argue that an image is normally associated with multiple contexts and the knowledge of these contexts can enhance the quantity of image meta-data available to the retrieval process.

The proposed context-aware retrieval system, capable of annotating the metadata associated with an image in a semiautomatic way by inferring the related information from captured contexts as well as other text-based or content-based information available for the image. The experimental results show that the proposed system outperforms conventional image retrieval systems.

2. Related Works

Context-aware computing is a paradigm associated with both mobile computing and human-computer interaction, and is strongly related to those working in ubiquitous/ pervasive computing, where the context is a key in their efforts to disperse and enmesh computation into our lives.

The concept of context-awareness was introduced by Schill, *et al* where it was used as a basis for designing adaptive software in mobile environments [5]. Their work describes how context-aware software can adapt according to location of use, the collection of nearby people, hosts, and accessible devises, as well as changes over time. During the last decade, design of adaptive systems based on context information has been targeted by many researchers [6] [7]. Much of this work has focused on location as the most important factor for determining context information.

Brown and Jones described a context-aware information retrieval application in their work [8]. Based on their description, a typical context-aware retrieval application involves a mobile user whose context is changing, and the retrieved information depends on the context. A tourist guide is an example of such applications [9].

There also exist a number of standards for image metadata specification that support descriptions of some

context information. Important standards for metadata specification include Dublin Core [10], MPEG-7 [2], and CIDOC/CRM [11].

Dublin Core was originally developed for description of text objects found in libraries, archives and government. It has been extended to facilitate image material such as that found in museums. The 15 basic descriptive elements in Dublin Core are organized in 3 categories and include elements such as Title, Subject, Source, Creator, Contributor, Publisher, Date, format, and language.

MPEG-7 has been developed by the ISO/IEC Moving Picture Experts Group (MPEG), and provides a framework for definition of metadata to describe the content of media objects, principally audio-visual objects. Unlike Dublin Core, no specific elements are defined, rather the framework includes two basic components: a descriptor and a description scheme that can be used to describe aspects of the media data. These descriptors are envisioned and implemented as xml tags.

CIDOC/CRM (CIDOC Conceptual Reference Model) from the International Committee for Documentation (CIDOC) provides definitions and a formal structure for describing events, changing attributes and dynamic relationships associated with a resource. CIDOC CRM is focussed towards describing physical museum artefacts and real world events, but has some limited abilities to describe digital objects and particularly digital multimedia or audiovisual content.

3. Semantic Understanding Using Contexts

To see how context information can be used to support semantic-based image retrieval, we first introduce image contexts and then examine how they contribute to tighten the gap between the user's needs for semantic retrieval and shortcomings of the current content-based image retrieval techniques.

3.1 Image Contexts

Dey & Abowd defined *context* as "any information that can be used to characterize the situation of an entity" [6]. Here, the entity is first of all an image, but it may also be a user searching for images to be used in a specific situation. In other descriptions of context, the synonym "environment" is used to give an understanding of the concept.

An image may be used in a number of different situations or contexts, and each context may emphasize the content of the image differently, giving a specific semantic understanding of what the image is about. An image may thus have a number of associated views, where each view reflects a specific focus or user interest in the image. For example, an image of Colosseum (an ancient Roman amphitheater) in a report on Roman architecture may be viewed as an example of an amphitheatre, while in the collection of holiday memories the same image shows one of the famous sights in Rome. As another example, consider the image of Fig. 1, which is a landscape of Gwangju Institute of science and Technology (GIST) in a winter day. While someone may be focusing on the administration building in the center of image, someone else may focus on the department of Mechatronics (i.e. the building in the left side of the picture), and others may consider the image as a general landscape of GIST in winter.

Fig. 1 A landscape of GIST in winter

Generally, contexts of different types may be useful for understanding the semantics of images. Here, we consider three general kinds of contexts: (i) spatial context; (ii) temporal context; and (iii) social context, as done by Davis *et al* in their paper [12]. Spatial context is related to the location where the picture is being taken. Temporal context can tell us for example the time when the picture is being taken, and finally social context determines to what objects a picture may be related or which objects may be included in a picture.

In addition to these general contexts, we use some more complicated contexts in our work as well, although they might be a direct or inferred combination of the abovementioned contexts. Two possible examples of such contexts may be context of origin (spatial context) and context of usage (social context), where the former refers to the situation when the image was first created, including relationships between the image and (i) real-world objects (such as persons, buildings), (ii) places (such as a city, street, valley), and (iii) events (such as festivals, sporting events, natural disasters), and the latter holds information about the environment where an image (or an instance of an image) is used, for instance an image collection or a report where the image is used as illustration [13].

3.2 Semantic and Sensory Gaps in CBIR

As mentioned before, there are still some major problems in the field of CBIR. In general, most of these problems can be divided into two major categories known as "semantic gap" and "sensory gap" [3] [4].

The semantic gap is described as the gap between the high-level semantic descriptions humans ascribe to images and the low-level features that machines can automatically parse. For example, a picture of a man tossing a red ball to a dog would be "seen" by a vision system as a series of moving color regions. The relationship between the man, the dog, the location where the ball is being thrown and the significance of this event to the person taking the picture are all gone. The most common means of attempting to solve this kind of problem are by adding captions or annotations to images. This however, is a costly and tedious process that requires many hours of effort, tweaking of machine algorithms, and careful watch over vocabulary and content to make sure that the images are tagged correctly.

In addition, most previous work in image annotation is done long after the image has been created, where it is most difficult to extract useful information about the image.

The sensory gap is described as the gap between an object and the computer's ability to sense and describe that object. For example, for some computational systems a "car" ceases to be a "car" if there is a tree in front of it, effectively dividing the car in two from the machine's perspective. One general idea for resolving this problem is that the domain and world knowledge should be explicitly built into the system. Knowledge that describes physical laws, laws about how objects behave and how people perceive them, and other supporting rules and categories are incorporated into the system in the hope of improving recognizers and helping machines to bridge the sensory gap. However, this type of knowledge-based approach has only really been viable for highly constrained, controlled, and regularized domains such as industrial automation applications.

Regarding these problems, it is easily seen that the context information can resolve them considerably if utilized in an efficient way. To do this, we propose a system whose details are discussed in the next section.

4. Proposed System

Based on the discussion of the previous section, in our proposed system, we focus on the image contexts, at the time of capturing as well as inferred contexts, to bridge the sensory and semantic gaps in image content annotation and retrieval. Figure 2 shows the framework of the proposed context-aware image retrieval system. As shown in this figure, the system consists of image retrieval engine, context capturing, annotation, context management, query processing, and context ranking subsystems.

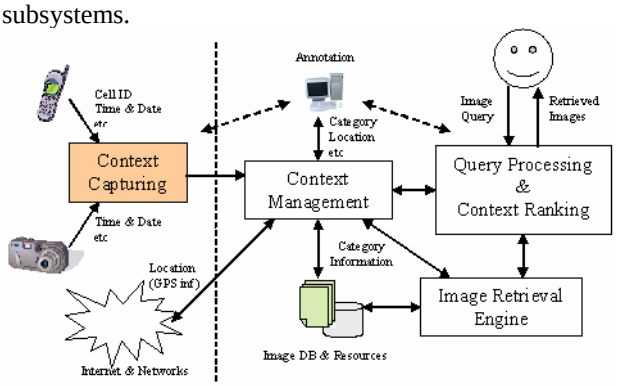

Fig. 2 The framework of the proposed system

4.1 Image Retrieval Engine

This engine is responsible for finding similar images from image collections based on the contents of the image query or in a text-based sense using key indices (not necessarily in a semantic form) associated to the images.

4.2 Context Capturing and Annotation

To support context-aware image management, we need methods for capturing context information of different types. For simplicity and without loss of generality, we suppose that the images taken by cellular phones or digital cameras include some basic contextual information (such as timestamp and cellular id). We then use these kinds of information as well as the information from similar images in other collections to infer new contexts for images which can be then used for retrieval purpose. So, we seek to support as much automatic capturing of inferred context information as possible, but realize that manual annotation may also be necessary.

Let's clarify the discussion through an example. Suppose that some pictures are taking by some persons during their visits from a temple in South Korea, where they use their cellular phone for this purpose. So, the images consisting of the following contexts: cell ID, date, and time. Using cell ID, we can then get GPS information of the location where the pictures are taken. In a later step, after uploading the images on the web, we can search for similar images on the collections based on their contents (as well as possible text indices) using the retrieval engine. We can then categorize the images based on the following rule:

$$(A \cap B) \cup C \qquad (1)$$

where A is the set of the all captured images with similar user (cell ID), B is the set of all images taken in a time slot), and C is the set of the top-ranked retrieved images.

The final metadata for images which satisfied the above conditions can be user, timestamp, category (i.e. an ancient temple), location, and similar existing text indices. Despite this, some manual annotation or user's feedback may be needed to prevent from some miscategorization.

4.3 Context Management

Image context information must be further analyzed to compute image descriptors that are used for both image retrieval and indexing. Context descriptors must be managed so that image retrieval systems can use context to enhance the knowledge of images and thereby support semantic-based image retrieval.

4.4 Query Processing and Context Ranking

Since context information is combined with text-based and content-based search, query specification and processing should support a combination of these elements. This subsystem is responsible for supporting such a combination. Also for final ranking of the result sets, this subsystem utilizes context information as an additional parameter.

5. Experimental Results

To show the effectiveness of utilizing context information in image retrieval systems, we developed a simple prototype of the proposed framework, where the retrieval engine is working based on the information of color and texture features in images, as well as text key indices. The color information is extracted using color histogram and the texture information is extracted using Gabor filtering technique [14] [15].

In an online process, the user first submits a text query (in the form of some keyword) and/or a sample images and then the system searches for similar images in different collections. Annotation and indexing is done in an offline process during context capturing as described before.

Here, we used four different kinds of collections to evaluate our system. The first collection consists of 500 images belonging to a CBIR project at the department of computer science and engineering in university of Washington [16]. Some typical images of the test data collection are illustrated in Fig. 3. These images include some general (non-semantic-form) key indices. For instance, in lower left images, the indices are *elk*, *tree*, *grass*, and so on.

We also select another group of images (about 300 images) from the same collection, however without any associated indices. The third collection consisting of 100 images is related to some ancient civilizations in Europe (e.g. Acropolis) and Asia (e.g. Persepolis) collected from local databases (Fig. 4). The indexing mechanism for the images of this collection has been done in a semantic-based method. For instance, *Ancient Civilization* → *Greece* → *Athena* → *Acropolis* → *Ephesus* → *Roman Buildings*.

Finally, the images in the last category has been supposed to be taken by some cellular phones and digital cameras for which we have all or some of the basic context information such as timestamp, cell ID, and GPS information. The total number of images in this collection is about 100.

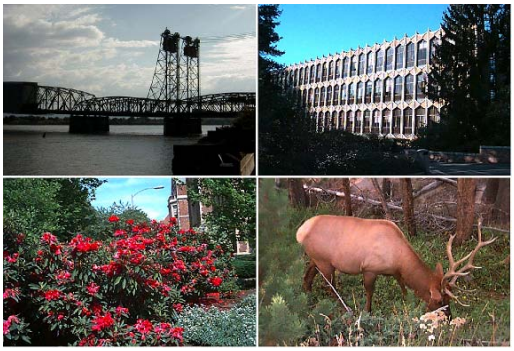

Fig. 3 Some typical images from the test collection

In the offline process, Annotating and indexing processes was first done for all images from different collections and then we examined through some experiments how context information improved the retrieval accuracy. For evaluating retrieval accuracy, in this work, we use *Hitin_M* parameter defined as:

$$HitIn\_M\ (Q) = N_R / M \qquad (2)$$

Where Q is the query submitted by the user and N_R is the number of relevant images in M top most candidates of retrieved images.

Experiment I:

- Scenario: I'm now in **Persepolis** (ancient capital of Iran) and want to make a report about **ancient civilizations similar** to Persepolis in **architecture**.
- Query: Find images related to ancient civilizations similar to Persepolis?
- Search Method:
  1. Text-Based: Ancient civilization and Persepolis
  2. Text-Based: Ancient civilization or Persepolis
  3. Content-Based: Query image of Fig. 4(a).
  4. Content-Based: Query image of Fig. 4(b).

(a) (b)

Fig. 4 (a) a photo shot from Persepolis
(b) a photo shot from Acropolis

  5. After annotation based on the context information.
     Location: Persepolis or Cell. ID → GPS inf.
     Category: Ancient civilizations
     Criterion: Similar Architecture (content)

- Result: HitinM for different search methods (Table 1)

Table 1. Hitin_M values for the experiment I

M	10	30	50
S. 1	20%	17%	16%
S. 2	60%	47%	36%
S. 3	50%	43%	34%
S. 4	60%	53%	36%
S. 5	80%	73%	70%

Experiment II:

- Scenario: I'm in **Montana**, US at a Hotel near **Rocky mountain Range**. I want to search for **national parks** in this area, not farther than 50km from my current position?
- Query: Find images of national parks in Montana not far from my current position?
- Search Method:
  1. Text-Based: Montana and National Park
  2. Content-Based: Query image of Fig. 5.

Fig. 5 A picture from Rocky mountain range

  3. After annotation based on the context information.
     Location: Hotel Name or Cell ID → GPS inf.
     Category: National Park
     Criterion: Not far from my current location
     Consideration:
     - If available, select pictures from photo galleries of national parks in US (specifically those which are in Montana or in its vicinity).
     - Use contents of retrieved images in next iterations, if the image is related to a national park in this vicinity (see Fig. 6 as an example).

Fig. 6 A picture from a national park in Montana

- Result: HitinM for different search methods (Table 2)

Table 2. Hitin_M values for the experiment II

M	10	30	50
S. 1	40%	33%	30%
S. 2	20%	17%	16%
S. 3	70%	70%	66%

Experiment III:
In this experiment, we generated 30 queries similar to the ones discussed in previous experiments and then calculated the retrieval accuracy for the system as follows:

$$Accuracy = \frac{\sum_{i=1}^{30} HitIn\_M(Q)}{30} \quad (3)$$

The result for different schemes is shown in Table 3.

Table 3. Accuracy values for the experiment III

M	10	30	50
Text-Based	50%	46%	44%
Content-Based	40%	35%	32%
Proposed Method	70%	67%	64%

With respect to the above experiments, we can see that the context-based image retrieval outperform the text-based and content-based image retrieval significantly.

6. Conclusion

In this paper, we proposed a context-aware image retrieval system, which use the context information associated to an image to infer a kind of metadata for the image which in turn improved the retrieval accuracy. We also examined that the knowledge of the context information can enhance the quantity of semantic understanding of an image. Experimental results show that the proposed system can significantly increase the retrieval accuracy in conventional image retrieval systems. In a future activity, we expect to increase the annotation (and as a result indexing) capability by considering user relevance feedback in the system.

Acknowledgement

This work was supported in part by GIST, in part by IITA, in part by MIC through RBRC, and in part by MOE through BK21 project.

References

[1] R.C. Veltkamp and M. Tanase, Content-Based Image Retrieval Systems: A Survey, A revised and extended version of Report UU-CS-2000-34, Computer Based Learning Unit, Universiteit Utrecht, 2000.

[2] MPEG-7 Standard: Overview, 2003. www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[3] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, Content-Based Image Retrieval at the End of the Early Years, IEEE Trans. on Pattern Analysis & Machine Intelligence, Vol. 22, pp. 1349-1380, 2000.

[4] R. Zhao and W. Grosky, Bridging the Semantic Gap in Image Retrieval, in Distributed Multimedia Databases: Techniques and Applications, T.K. Shih (Ed.), Idea Group Publishing, Hershey, Pennsylvania, 2001.

[5] A. Schill, N. Adams, and R. Want, Context-Aware Computing Applications, in Proc. Of the Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, December 1994.

[6] A. Dey and G. Abowd, Towards a Better Understanding of Context and Context-Awareness, in Workshop on What, Who, Where, When, and How of Context-Awareness, Conference on Human Factors in Computer Systems, 2000.

[7] G. Chen and D.A. Kotz, A survey of context-aware mobile computing research, Tech. Rep. TR2000-381, Dartmouth, November 2000.

[8] P.J. Brown and G.J.F. Jones, Context-aware Retrieval: Exploring a New Environment for Information Retrieval and Information Filtering, Personal and Ubiquitous Computing, 2001.

[9] S. Long, R. Kooper, G.D. Abowd, and C.G. Atkenson, Rapid prototyping of mobile context-aware applications: the Cyberguide case study. In Proc. of the Second Annual International Conference on Mobile Computing and Networking, White Plains, NY, November, 1996.

[10] Dublin Core Working Group, 2001-2004. www.dublincore.com

[11] CIDOC Conceptual Reference Model (CRM), 2004. http://cidoc.ics.forth.gr/

[12] M. Davis, S. King, N. Good, and R. Sarvas From Context to Content: Leveraging Context to Infer Media Metadata. Multimedia 2004, New York, NY, USA. ACM Press 2004.

[13] R. Karlsen and J. Nordbotten, CAIM: Context Aware Image Management, Project Proposal, Dept. of Information and Media Science, University of Bergen, 2005.

[14] B.S. Manjunath and W.Y. Ma, Texture Features for Browsing and Retrieval Image Data, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, 1996.

[15] M.R. Hejazi, Y. Ho, A Rotation-Invariant Content-Based Image Retrieval based on Color Histogram, Texture Feature, and Radon Transform Parameters, Proc of IT International Student Fair 2005, Suwon, South Korea, August 2005.

[16] www.cs.washington.edu/research/imagedatabase/