An Analysis of Image Cognition Structure for Metadata Type Image Retrieval

Toru FUKUMOTO, Kanji AKAHORI
Department of Human System Science, Graduate School of Decision Science and Technology, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo, JAPAN
+81-3-5734-3233
fukumoto@ak.cradle.titech.ac.jp

Abstract

Nowadays, a large amount of digital images are being stored worldwide in the Internet. As an educational means, images stored in it have big potential. And it is so rapidly expanding and becomes so complicated that the ways to retrieve images effectively are getting more difficult. As a Digital Disc Camcorder or movie delivery via the Internet, not only still images but also motion pictures will be stored and retrieved seamlessly.

We considered the structure of metadata suitable for storing and retrieving seamlessly from one system both them. We examined what part in them people pay attention to and the difference between the parts. The order of keyword attached to images is proper noun, situation, event, action of main object, action without the main object or background, and impression. And the subject describes the keyword first which has an affective impact in an image. And we may say that the structure of metadata needs to be prepared for three patterns: still image, motion picture like still, and motion picture. Appling these rules, database administrators efficiently attach keywords to images. At retrieval, the system will show better results. And users store and retrieve suitable data for each media type or content.

1.Introduction

Over the years large amounts of computer-aided images are being stored in the Internet owing to widely available digital recording devices. As a Digital Disc Camcorder or movie delivery via the Internet, not only still images but also motion pictures will be stored and retrieved seamlessly.

There are 3 kinds of image database: feature type, sensitiveness type, and metadata type. Our concern is that at one system will seamlessly be stored and retrieved both still images and motion pictures. So feature type and sensitiveness type are not suitable for retrieving motion pictures. Because it is difficult to extract color histograms or shape from them (MPEG-7,2001). This is why we focus on metadata type in this paper. Of metadata type, first, database creators define the structure or the framework of metadata. Second, database administrators attach metadata to images according to it in the database. Third, a

retriever specifies texts as a key to the database. Finally the database system searches images using the metadata which is given by the administrator and also using the texts which are keyed in by the retriever. Examples of metadata are keywords, texts, classification items and so on.

By the way, we may note, in passing, the indexing of image. In the area of art documentation, there are some trials to describe picture content (UEDA,1997), 4W method, PDL method, and Iconography. But they mainly depend on the painter's intention or bibliography of a picture itself. So they are hard to apply image retrieval.

2. The Framework of Metadata

We have already dealt with the criteria of metadata/keyword (FUKUMOTO,2000). For storing and retrieving seamlessly at one system both still images and motion pictures, we are now concerned with the structure of metadata suitable for them. For this purpose, we examine what part in motion pictures or still images people pay attention to. If approved, this will be a guide on attaching metadata for database administrators. And is it different the parts between motion pictures and still images that people pay attention? If so, the structure of metadata attached to still images and to motion pictures should be different. This leads database users to store and retrieve contents suitable for each media type. In the next section the method of our experiment. In Section 4 the result is discussed. In Section 5 our conclusion is presented and the future work is discussed

3. The Procedure of Experiment

Subjects of the experiment are 28 undergraduate students. All are accustomed to search engine in the Internet. First of all, We divided 28 students randomly into two groups; Group A and Group B. Group A were showed the motion pictures first and the still images subsequently. Group B were showed the still images first and motion pictures subsequently. Next we let them to show still image or motion picture. The contents are 13 patterns of various types: news, soccer, baseball, drama, sight, and animation. And we prepared two or more contents of each type (news, soccer.) The motion picture consists of one scene, from a switch point of a scene to the next one, and the still image is extracted as key frame of the motion picture. The motion picture and corresponding still image are same contents for the previous reason. We let the motion picture repeat three times and still image show as much time as the motion picture corresponding to. Then we let them give keywords to both. They gave as many keywords to them as they liked. We did not limit the number of keywords.

4.The Result of Experiment

4.1 Details and Orders of Keywords

The details and orders of keywords have a tendency as follows. First, proper noun is described when a place or person is specified in a still image or motion picture, for instance, "Clinton" or "Akebono" who was a famous sumo player. Second, a situation is described, "baseball" or "Drama". Third, an event is described, "bat and pitch " or "meeting". Fourth, the action of main object in a still image or motion picture, "shoot" or "speaking". In one situation two or more events or actions may exist. In one situation "news", one event "goal in" and two action "pass" and "shoot" are described. Fifth, the background or the actions of objects except the main object are described, such as "full of audience" or "chest". Last, an impression or an

atmosphere from a still image or motion picture is described, “fine” or “comfortable”. This tendency, we call as Rule 1, explains at 47.7% of all descriptions. By the way, there is no significant difference in the details and orders of keywords between still images and motion pictures. That is to say, there is the same tendency towards them..

There is one more rule to extend the previous one. For an image or picture which has a peculiar event, the event is often described before a situation. In figure Ⅰ, there are some cases peculiar to the event “shoot”, the order is first “shoot”, second “soccer”. The keywords about impression or atmosphere are similar. In TableⅡ, there are other cases such that the impressionable word “cheers !” is described first because this subject is a fan of this team when we interviewed after. This rule, we call as Rule2, explains at 20.5% of all. In other words, the subject described the keyword first about which an affect have impact on. In Table Ⅰ, the order is generally “soccer”, ”shoot”, “goal”. According to this rule the order is “goal”, “soccer”, “shoot ”. It is the same as previous rule, and there is no significant difference in this rule between the still image and the motion picture.

Appling this rule, database administrators efficiently attach keywords to images in a database. So a computer as programmed this rule can support them. To put it the other way, if an image is attached keywords ordered with this rule, an affective object exists in the image. With this tendency, searching keywords prior which are attached first when inputted words by a retriever, the system shows better results. As a result, these two rules explain at 68.1%.

Situation	Event	Action	Impression
Soccer	Goal	Shoot	Fine !

Rule 1

Action	Situation	Event
Shoot	Soccer	Goal

Rule 2

Impression	Situation	Event
Fine !	Soccer	Goal

Rule 2

Table Ⅰ : Two Rules of Order

We may note to consider extracting the order of keyword from Rule 1. This means focusing attention on the details of keywords and the extension of Rule 1. This extension explains at 7.8%, and so Rule 1, Rule 2,and this extension explain at 75.96%. That is to say, to restrict only the content of keywords and to attach them to an image or a picture is effective for image retrieval.

4.2 Number of Keywords

We made sure of the difference between a still image and a motion picture in the points which subjects pay attention to. So we counted the number of keywords both a still image and a motion picture. We present the results in TableⅢ. The result shows that they attached at a motion picture more than or the same at a still image at significant difference of 1% or 5% level. They may be accustomed to motion pictures. They catch description from contents and attached keywords at a motion picture

more easily than at a still image. And we picked and counted only action words. We present the result in Table Ⅳ, at a motion picture more than a still image at significant difference of 1% or 5% level. They observed the actions of a motion picture more easily than a still image. By the way, Group A has no significant difference from Group B.



Figure Ⅰ: Examples of Contents

4.3 Variety of Keywords

We counted without redundancy the number of keywords from both a still image and a motion picture. With some contents they attached at a motion picture more than a still image at significant difference of 1% or 5% level. By the way, Group A has no significant difference from Group B. We shall concentrate on the contents at difference of 1% level. The common feature of these contents is that these motion pictures have little movement and resemble still images.

If a motion picture has little movement, subjects feel as though it were both a motion picture and a still image. There are keywords for still image and about moving object. To sum up, first, as showing a still image, a person attach keywords to an image, and next, as showing a motion picture, he/she attach keywords. From a view of attaching keywords, such a motion picture has the features of both a motion picture and a still image. On the contrary, in motion pictures which have a lot of movements, we observed that subjects concentrate at only moving objects. From these results we may say that a person looks at images or pictures in three ways: still image, motion picture like still, and motion picture.

5. Conclusion

In this paper, we considered the structure of metadata suitable for storing and retrieving both still images and motion pictures seamlessly at one system. We examined what part of motion pictures or still images people pay attention to. And we found that there is different aspect between motion pictures and still images that people pay attention. We may say that the structure of metadata needs to be prepared three patterns: still image, motion picture like still, and motion picture. This leads database users to store and retrieve suitable for each media type or content.

In the future, we hope that multimedia retrieval with metadata will be much easier for all people. So the teachers will be able to utilize more accurate images in education, and the students will be able to easily retrieve images from the Internet.

References

Yasushi KIYOKI et.al., “A Semantic Search Method and Its Learning Mechanism for Image Databases Based on a Mathematical Model of Meaning”, Trans. IEICE Japan D-Ⅱ Vol.J79-D-Ⅱ No.4, pp.509-519 (1996).

MPEG-7 Web Page: http://www.mpeg-7.com/ (accessed on March 1, 2001).

Shuichi UEDA, “Pictorial Database –Indexing and Retrieval Methods”, Journal of IPSJ,Vol38 No.5 pp.401-404 (1997)

Toru FUKUMOTO and Kanji AKAHORI, ” The Criteria and Evaluation of Metadata/Keywords in Image Retrieval”, Proceedings of the ICCE/ICCAI 2000, pp.542-546 (2000).