

Digitalisation of Film Material and Content-Oriented Retrieval

Young-Man Ko

Department of Library and Information Science, Sung Kyun Kwan University, Korea

Abstract

The moving image can be nowadays digitalised by new information technology and this can be found through the telecommunication networks such as Internet. Information services now aim to provide the content-oriented retrieval in a convenient way. This aim is also valid for the information service in the field of film materials. In order to fulfil this purpose the system must be able to recognise the image and sequence of films. Furthermore, new data types and data elements must be defined, crystallised and standardised. The following paper outlines the development of information work in film archives, concepts on the content-oriented information work and retrieval for moving image materials, definition of new data types and new data elements.

Keywords : Digital film, Moving image, Content retrieval, Data element, Data type.

1. Introduction

Due to the development of digital technology, it is quite possible to digitalise moving image material. However the significance of digital technology goes far beyond merely creating the digital image from the original materials. They have a greater impact on information work and retrieval in film and media archives.

To say that information services nowadays aim to provide the content-oriented retrieval in a convenient way, is true. This aim is also valid for the information services in the field of film or moving image materials. The Internet service "WWW" shows clearly that such a service is already being developed.

However the content-oriented information service must fulfil two basic requirements:

- There must be the possibility to recognise the image, cut or sequence. In order to fulfil this requirement it must be an automatic recognising system. The prototype, which recognises the "hard cut" has already been developed. The system for "soft cut", such as a complex camera moving/focussing is to be produced in the next few years.
- There must be the possibility to process information work in various ways. The pre-requisite for this process is the definition of the data types and the selection of the data elements.

The purpose of this paper is to discuss the definition of the data types and the level of the data elements in relation to the digitalisation of film materials and retrieval of film or film parts.

2. The needs to define the new data types and data elements of film materials

The complete film has almost been regarded as one documentation unit in the film and media archives, and thus the information work in the media archives has traditionally concerned merely filmographical processing, so far. The filmographical information includes generally formal elements of the complete film (Sul, 1996).

Film material can be digitalised by new digital technology and digitalised moving pictures can be found through the telecommunication networks such as Internet (Wyatt, 1996; Korea KODAK, 1996). Moreover many projects have started developing the tools which can recognise the frame, sequence, definite objects or sounds automatically (Süllow, 1996). In this case, the content-oriented access assume even greater importance than ever. As we know, such an access has been already realised in the field of text information service by the new methods of knowledge organisation and knowledge representation like hyper-linking and hyper-media.

The new development of digital technology for the moving images creates the additional needs to retrieve the film sequence by definite length, camera focusing, lighting etc. Consequently the information work for film materials is now confronted with this new challenge to crystallise the new data elements and to define the new data types (Ko, 1996). In addition to the traditional criteria the new data type must enable the information professionals or documentalists to process information work in order to access directly the definite content of the film materials and film parts (Srihari, 1994).

3. The development of information work in film and media archives

The information work in film and media archives has mostly concentrated on the formal development of processing. That is until now. Even though some media archives have tried for content-oriented processing, the complete film has served as one unit of information processing, e.g. thema, genre, abstract etc. The content-oriented processing has stood back out of the formal filmography.

The content-oriented processing for film materials was primarily developed in the information centre of broadcasting institutions because their programmes are very often in need of the definite sequences like the former speech of the president. In relation to this kind of information processing, some curriculum's on the analysis of the moving images, are

provided at several universities, e.g. Emporia State University in the USA. These subjects are primarily concerned with information/knowledge representation and information marketing/brokering in information science. These lectures can be supported by traditional techniques like foreword and backward running equipment or some digital technologies like CD-Rom, but the moving images are analysed by manual processing (see Figure 1). Thus such a processing can be realised only under the circumstances of the very high investment for personal cost and time.

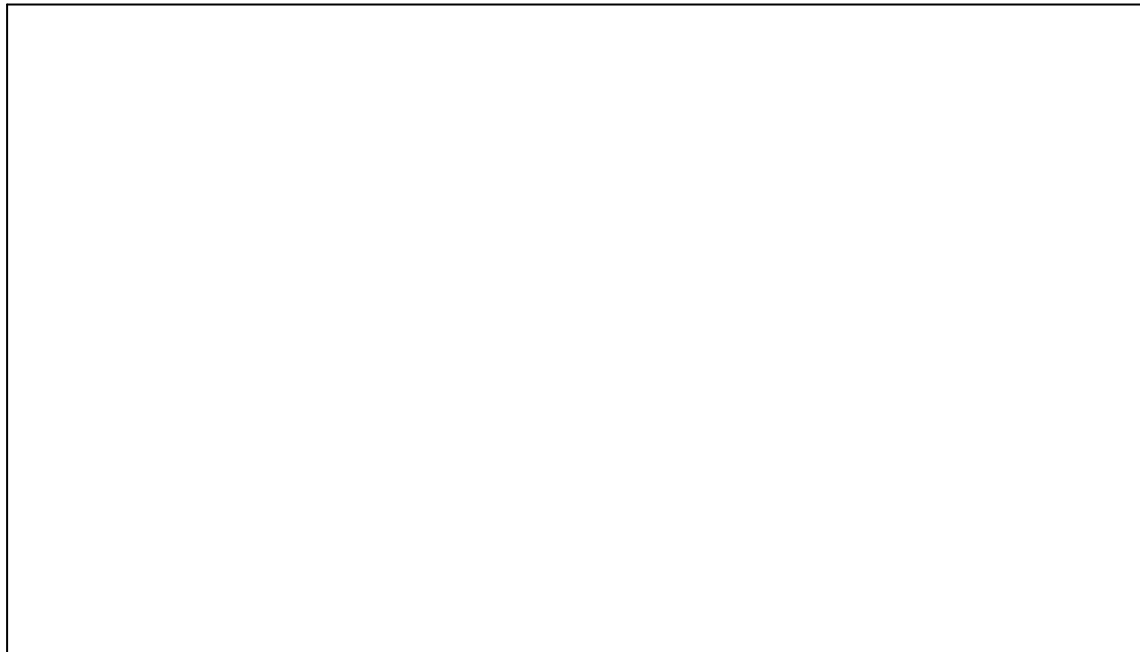


Figure 1: 20 seconds segment from 'Pretty Woman' (Kim, 1995).

Table 1: Development of information processing for moving image.

PROCESSING STAGE	Filmography	Content-Oriented Processing	Content-Oriented Communication	Marketing	Cost Reduction Effect
1. Manual Processing	o	x	x	x	x
2. Traditional Technology & Manual Processing	o	□	x	x	x
3. Digital Technology & Manual Processing	o	o	□	□	□
4. Digital Technology & Automatic Processing	o	o	o	o	o

(* o = possible □ = partly x = impossible)

As a result, discussions have recently started on the use of the new information technology and tools which can realise the content-oriented processing and retrieval - more exactly - cheaper and automatically than ever (Calson, 1993). So we can now classify the development stage of the information work in the film and moving image archives as follows: (see Table 1).

4. Concepts on the content-oriented information work and retrieval for film materials

As seen at the illustrated part in the bottom of Table 1, the fourth stage of information processing has not entirely been realised. It is possible however and many approaches are now in progress. According to Süllow (1996), diverse approaches to realise the fourth stage can be summarised in five main streams as follows:

- (1) Automatic dissection of film in shares of sequence,
- (2) Automatic recognising of the moving images by specialised knowledge base system,
- (3) Automatic indexing of the image information by the use of concomitant information like commentaries or captions,
- (4) Automatic comparison with moving images by the use of browsing tools like explicate method (to compare to similar figures) and implicate method (supported by questionnaires and references of the qualified retrieval system),
- (5) Detailed description on the contents with regard to their atmospheres and associations, but this can't be automated. This needs additional implementation of the suitable fulltext database.

One of the projects which can realise the fourth stage in the foreseeable future is the Project AMPHORE (Audio-Visual Media Platform for the Highlighting, Organising and Retrieval of Entities). The project is based on the concepts 1, 3, 4 and 5, because the building of such a knowledge base system is very difficult and the system 2 can't recognise the atmospheres and associations of the concerned scenes anyway. The project AMPHORE started on the basis of co-operation between IWF (Institut für den Wissenschaftlichen Film) Göttingen and IPSI (Institut für Integrierte Publikations - und Informationssysteme) of GMD in Darmstadt on Autumn 1994.

"AMPHORE is a client-server system for the documentation of moving image material. The server mainly is formed by a fulltext database with SGML capabilities while the clients are PC working places equipped with software for documentation and retrieval of movie and/or movie parts. In AMPHORE, the

complete film material is provided in digital form and thus can be used for content-oriented documentation and retrieval. This enables the documentalist to build very detailed indexes allowing access by sequences or even by shot. The film descriptions are based upon a syntactical thesaurus - controlled indexing which reflects the film's diverse action string and levels." (Süllow, 1996).

5. Definition of new data types and new data elements

The data types and data elements which has been defined for the information processing and retrieval can be classified exemplary by their development stages. The media archives which the manual information processing is realised in have generally very simple data types: formal filmography and filmography in content. Formal filmography includes, according to KBS, MBC and MaeKyung, call no., location, broadcasting datum, length of film, kind of materials, produces, director, step, casting etc. Filmography in content consist of title, genre, original work and author, scenario and writer, abstract, keyword, classification by theme etc.

As the analysis of visual imagery "Pretty Woman" shows, the information work in which manual processing and the use of techniques belong together has content-oriented data types and data elements, namely camera shot (long shot, medium shot, close up, zooms), camera and subject movement (dolling, back screen projection) and sounds (dialogue, sound effects, music, narration, silence) etc.

Finally the example AMPHORE which is going to realise the full automatic processing and retrieval provides more content-oriented and more detailed data types defined for information processing and each data type includes the following data elements.

- (1) data type **Document** : formal and content filmography
- (2) data type **Film Data** : reference information to the files in which each digitalised film is stored separately.
- (3) data type **Sound Text** : concomitant information like commentary, dialogue or captions.
- (4) data type **Attendant** : comment or review, scenario, scientific paper on the concerned films
- (5) data type **Sequence** : different kind elements by frame and sequence, e.g. technical information, optical information, thesaurus for description of definite sequence.

Besides, the AMPHORE provides retrieval system which consists of following types of operating tools (Süllow, 1996):

- (1) operator **Descriptor** : It allows the user to retrieve the definite sequence by

questionnaires with 'subject-verb-object' structure. The retrieval is realised by compulsory descriptor (subject) and selective descriptor (verb and object), e.g. [lion], [plane, fly] or [fox, hunt, rabbit].

- (2) operator **Fulltext** : It supports the user to retrieve sound information or attendant information by stop-word method.
- (3) operator **Document** : It consists of traditional filmography.
- (4) operator **Technique** : With the help of this operator the user can find the definite sequence or frame according to their technical information needs.

These four operators can be used at the same time by combination with "and" and "or" logics.

6. Solution, perspective and conclusion

(1) Communication-oriented information work

In order to realise the communication-oriented information processing, the system must fulfil the following basic requirement

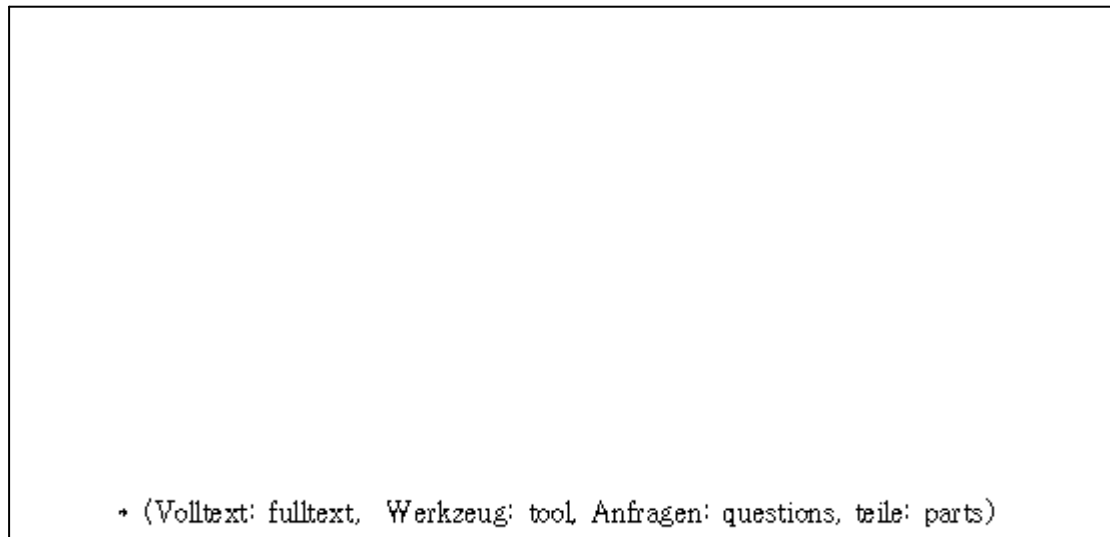


Figure 2: Client - Server - Architecture of AMPHORE system (Süllow, 1996)

First of all, the success of the film databank consist in the storage capacity and precise retrieval. The problem of storage could be solved by MPEG-Coding or Motion-JPEG and a project for the development of "High Level MPEG" is now in preparing of ETRL Korea

(Lee, 1996). There are different methods for the loading of digitalised materials to the client-system: storing the whole films in a file of server-system and retrieval like the Internet service "WWW", storing the whole films in a file of server-system and retrieval after loading of the concerned film to the client-system, storing each film in their own file separately and retrieval. The selection of the methods depends on certainty, preciseness and rapidity of the retrieval and the optical quality of the pictures.

Secondly, the system must have "Client - Server - Architecture". For example the AMPHORE provides this Architecture (see Figure 2).

Thirdly, there must be the qualified browsing tool and the suitable thesaurus-structure which permit the syntactic retrieval and protect the intraindexer effect (Enser, 1995).

(2) Marketing-oriented information service in film and media archives

Information marketing or brokering of the complete film material is nowadays very popular either in the foreign or in the domestic markets. It is also expected that the market of the content-oriented information retrieval will be of important as soon as this service is provided, because there are many potential clients, e.g. moviemaker (producer, director, scenario writer, scriptor, actor and actress, film engineer, technician, film studio, cinema house, post production, video production), research institute, university (professor, student), broadcasting institutions, press agents, advertising companies, information provider and information distributor like Internet and Nownuri, film fan (above all film fans on Internet), etc. The increasing number of Internet users and the statistical yearbook of Council of Europe (1996) show obviously how great the potential market of the content-oriented information service for film material is.

The film and the moving image consist of the picture which magic, orientation, projection/identification, knowledge and sensuality exist in (Schuck-Wersig, 1992) and technology. The new digital technology has created diverse value-added effects, e.g. comparative, inherent, agglomerate or integrated effects in the field of the text information service (Kulen, 1995).

Furthermore, it is expected to create the esthetical value-added effect with the digitalisation of the moving images and the content-oriented information retrieval. This is the new challenge of the new age and this is a mission for us to create the new value of our work. So there are many things we must develop, crystallise and standardise in future.

References

- Carlson, C : Perspectives of a Hypermedia Film Sequence Database, In: Innovative Medienanwendungen, Beiträge der IAMS/GMW - Tagung, Juni 1993, Göttingen, Median in der Wissenschaft, Vol. 2, Göttingen, 1993, pp. 92-97.
- Council of Europe : Europäische Audiovisuelle Informationsstelle, Statistisches Jahrbuch 96, Nomos Verlag, Baden-Baden, 1996.
- Enser, P. G. B. : Pictorial Information Retrieval, In: Journal of Documentation, Vol. 51, 1995 (2), pp. 126-170.
- Jo, Hee Moon : Informatization of Film Materials, Workshop paper at ETRI, Yuseong, 30, May 1996 (in Korean).
- Kim, Jin Mook : Visual Imagery Analysis: "Pretty Woman", Seminar Paper at the Emporia State University, 1995.
- Ko, Young Man : Digital library and Database of Film Materials, Workshop paper at ETRI, Yuseong, 30, May 1996 (in Korean).
- Korea KODAK : Digitalisation of Visual Image by CINEON, Workshop paper at ETRI, Yuseong, 30, May 1996 (in Korean).
- Kuhlen, R. : Informationsmarkt, Konstanz Uni, Verlag, Konstanz, 1995.
- Lee, Ui Taeg : MPEG 2 Coding, Workshop paper at ETRI, Yuseong, 30, May 1996 (in Korean).
- Schuck-Wersig, P. : Expedition zum Bild, Beiträge zur Analyse des kulturellen Stellsenwert von Bildern, Peter Lang, Frankfurt a.M., Berlin, Bern, New York, Paris, Wien, 1992.
- Srihari, R. K. : Automatic Indexing and Content-Based Retrieval of Captioned Images, In: IEEE Computer, 1995 (Sep), pp. 49-56.
- Sul, Ki Whan : Direction for the Database of Film Materials in Korea, Workshop paper at ETRI, Yuseong, 30, May 1996, (in Korean).
- Sullow, K. : AMPHORE - Ein Arbeitsplatz zur Filmdokumentation, In: Nachrichten für Dokumentation, Vol. 47, 1996 (2), pp. 67-74.
- Wyatt, R. B. : Welcome to Digital Cinema Today, <http://tech-head.com/cinema.htm> (July, 1996).