MPEG-7 CAMERA

T. Ebrahimi, Y. Abdeljaoued, R. M. Figueras i Ventura, and O. Divorra Escoda

Swiss Federal Institute of Technology (EPFL)
Signal Processing Laboratoy (LTS)
CH-1015 Lausanne, Switzerland
WWW home page: http://ltswww.epfl.ch, FAX: +41-21-693-7600
{touradj.ebrahimi,yousri.abdeljaoued}@epfl.ch

ABSTRACT

MPEG-7 camera extends the capabilities of conventional cameras by analyzing its scene in order to generate a content-based description according to the recently approved MPEG-7 standard. This gives to the camera a large variety of current and potential applications, such as surveillance, augmented reality, and virtual display. This paper provides an overview of what is meant by an MPEG-7 camera, discusses the above mentioned applications, and provides an implementation example of such a camera using existing hardware products.

1. INTRODUCTION

Recent advances in the fields of networking, data processing, storage, and acquisition are driving the creation of large amounts of audio-visual data. Such data is often consumed directly by the user in its original form (modality). In this case, the standards MPEG-1 and MPEG-2, based on a sample-based representation, can be used to efficiently and compactly represent the audio-visual information. In order to allow interactivity with the audio-visual data, an object-based representation, as adopted in MPEG-4, was developed to be used in a broad range of multimedia applications. There is however an increasing number of applications, where the audio-visual data is further processed in order to be retrieved, filtered, and identified by both users and computational systems. For example, in a surveillance application, the scene is analyzed with the goal to detecting alarm situations.

Depending on the targeted application, the audio-visual data is represented by a specific content-based description. In order to avoid transmission or storage of irrelevant information present in the audio-visual data, we propose to perform the extraction of the content-based description at the acquisition stage, i.e. at the camera. Furthermore, the use of the standard MPEG-7 for the content-based description of the audio-visual data allows more flexibility and interoperability in the design of complex and powerful multimedia systems. The above approach also fits adequately the trend from data to information and from information to knowledge processing. An MPEG-7 camera can been seen as an interface from data to knowledge.

In this paper, we present some results obtained after a study motivated by the advantages of processing audio-visual data at camera level. To do so, we report a working implementation of an MPEG-7 camera, developed by augmenting a conventional camera with data processing capabilities. Such a camera delivers as output a content-based description of the scene according to the MPEG-7 standard.

2. WHAT IS A CAMERA?

Cameras are among common devices to capture audio-visual information from the real world. A camera can be seen as a converter transforming an audio-visual data (coming from a real scene) to an electric signal (camera output). The electric signal can be stored or transmitted and would need another converter (for instance a display or a printer) in order to reconvert (render) it back to its original form (modality) hopefully representing the original scene as closely as possible. Many efforts have been made in order to improve the efficiency of the above process. Among them one can mention progress in sensor technology, transmission, storage, display and print.

Early cameras performed all the processing in the analog world where the actual degree of intensity of the output signal would represent the actual information. With the progress of digital technology and fast and cheap digital to analog conversion, a new generation of cameras known as digital start appearing. This opened the door to a plethora of new innovations and applications. Because of the large volume required to represent audio-visual data, compression technologies were incorporated inside cameras, allowing a more efficient I/O. Simple processing such as automatic gain control, jitter cancellation, and picture enhancement also started to appear in both analog and digital cam-

eras.

One important feature in most consumer and professional cameras is the format of their output. In fact, for obvious reasons, the output of various cameras produced by different manufacturers follow a same standardized format in order to allow efficient use of the latter in the largest possible set of applications. A standardized output is therefore one of the key issues in cameras. Based on the above considerations, cameras with outputs compliant with standard compression algorithms have appeared. Among the most popular, one can mention JPEG still image cameras, as well as MPEG-1 and MPEG-2 cameras [1].

More recently, an MPEG-4 camera was introduced to the market which aimed at lower-end and consumer applications in streaming of video over the Internet [2]. Although, this camera has many merits, here we would like to propose an extended definition of an MPEG-4 camera beyond this specific product and its targeted application.

2.1. MPEG-4 camera

As explained earlier, MPEG-4 coding is based on an object-based approach to data representation. This means that an MPEG-4 camera could select specific objects in an audio-visual scene (AVO) and output only those selected , disregarding the rest. More sophisticated processing, such as various spatio-temporal processing of AVOs, or even insertion of other natural AVOs can also be part of the capabilities of such an MPEG-4 camera. An MPEG-4 camera, thanks to its object-based representation, is the first camera which does not reproduce an entire scene from the real world, but rather a device capable of outputting specific audio-visual objects in a very flexible fashion, not necessarily reflecting a real scene. Such a camera would trigger efficient implementations of many applications beyond those achievable with today's conventional cameras. Such a discussion however goes beyond the scope of this paper. We would just mention that by our definition, generation and coding of synthetic AVOs, although possible by MPEG-4 standard, is not considered as part of the operations of an MPEG-4 camera, but a later stage.

2.2. MPEG-7 camera

Following the same philosophy, an MPEG-7 camera would obviously convert an audio-visual scene into a content-based representation which is the basis upon which MPEG-7 standard is built. In addition to features of an MPEG-4 camera, (thanks to the fact that an object-based representation is a sub-set of a content-based representation), an MPEG-7 camera can offer a number of extended functionalities. One of the most important features of an MPEG-7 camera resides in the fact that it is independent of the modality of the captured data. In all other cameras mentioned above, the output representation is intimately linked to the modality of the original scene captured by the camera. For example, a scene captured by a JPEG or MPEG-2 camera is represented and therefore can be directly reproduced as an approximation of the same scene. An MPEG-4 camera can select specific objects in that same scene, but it will always represent objects similar to those captured by the camera, if no other sophisticated processing is applied. An MPEG-7 camera however would be capable of converting the scene into a description which could be later used or converted to more than its original form.

Another important feature of an MPEG-7 camera is the elegant and relatively easy way to replace or extend the conventional sensors (audio-visual) in a camera by others. Pushing this approach far enough, it is not even necessary to have a camera with optical sensors in order to capture and to display the movement of a person, as a motion detector could suffice. A multi-modal camera can be easily realized by multiplexing the modality independent descriptions coming from each sensor in a rather straightforward manner. It is obvious that an MPEG-7 camera in the way described above could require some degree of sophisticated processing inside of the camera itself. Such cameras are usually referred to as smart cameras.

In short, an MPEG-7 camera can be seen as a smart camera with an an MPEG-7 compliant format as output. It is important to mention that an MPEG-7 camera does not necessarily exclude integration with a more conventional camera, nor requires automatic and complex processing. A simple MPEG-7 camera could be a conventional composite camera enhanced with MPEG-7 compliant meta-data about the output (either entered as text by a user or extracted from camera parameters), adequately multiplexed with the latter.

The camera used within this work, has been developed by the company FAST*COM* Technology. It integrates in a single device the hardware for capturing visual information, image processing, and network connection. It consists mainly of a digital signal processor (DSP), a high dynamic CMOS image sensor (120 dB), a flash memory, and different RS-232 and RS-485 interfaces for interactions between the camera and outside (e.g. a PC).

2.3. MPEG-7 standard

The Moving Picture Experts Group (MPEG) has initiated a new work item known as MPEG-7: Multimedia Content Description Standard [3]. The goal of this work is to standardize content-based description for multimedia data that can enable applications such as multimedia information search, filtering, management and processing.

MPEG-7 specifies a standard set of Descriptors (D) for the representation of a feature, a set of pre-defined structures of Descriptors and their relationships called Description Schemes (DS), a Description Definition Language (DDL)

that defines description schemes, and a coded representations of a description for efficient storage, streaming and fast access.

Because of the popularity of the Extensible Markup Language (XML), the availability of tools, and the common goals, the XML Schema Language has been adopted as the MPEG-7 DDL. Since the XML Schema Language has not been designed specifically for audio-visual content, certain extensions are necessary in order to satisfy all of the MPEG-7 DDL requirements. These extensions are currently being discussed through liaison activities between MPEG and the World Wide Web Consortium (W3C), the group standardizing XML.

2.4. MPEG-7 camera structure

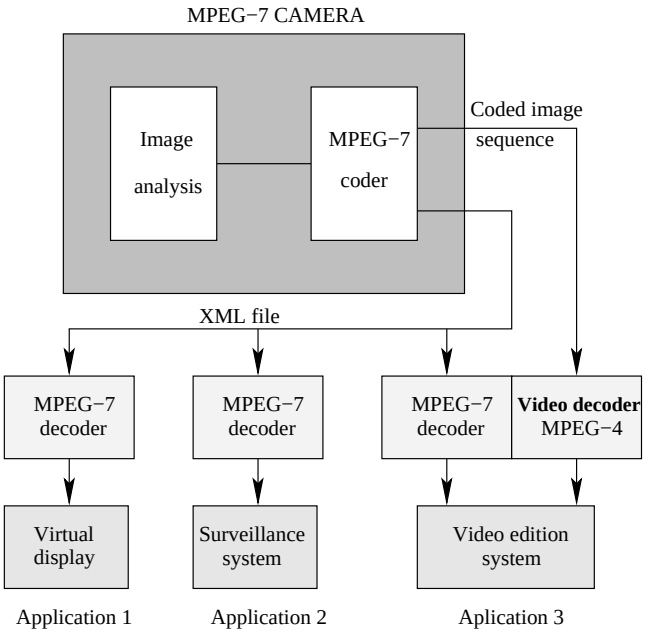

Fig. 1. Structure of the MPEG-7 camera.

A typical processing chain of the MPEG-7 camera is shown in Fig. 1. It includes the image analysis block which consists of a set of algorithms for change detection, feature extraction, and object tracking. One of the most important criteria by the selection of these algorithms is their low-complexity. This makes the implementation of the algorithms on the DSP feasible. Of course, only the features relevant to the targeted applications are extracted from the audio-visual data. Then, these extracted features are used for the instantiation of MPEG-7 description schemes in the MPEG-7 coder block. These MPEG-7 descriptions are coded either as binary or textual and sent to an MPEG-7 compliant device. In some applications like video editing systems, the MPEG-7 description is associated with the audio-visual data, represented for example by the MPEG-4 standard. At the application side, an MPEG-7 decoder is used to decode the MPEG-7 descriptions. Depending on the application, this description is consumed in a specific way.

3. EXAMPLES OF APPLICATIONS

A number of potential applications were identified by using the MPEG-7 camera. For each of these applications the related MPEG-7 DS's are selected [4]. Then, the algorithms for the extraction of the features, needed for the instantiation of the selected MPEG-7 DS's, are implemented on the MPEG-7 camera.

In a surveillance scenario, the delivered MPEG-7 *Moving Region DS* can be used to recognize changes in a given scenery. In the case of traffic monitoring, the MPEG-7 *Motion Trajectory D* [5], which is part of the *Moving Region DS*, is used to detect alarm situation such as traffic jam and accidents.

Another interesting application is augmented reality, where a virtual object is added to a real scene. Tracking the position of the user in an augmented reality system is one of the important issues for virtual object insertion. According to the estimated new position of the user, the visualization system renders a new view of the scene in a way that the user has the impression that the virtual object is part of that scene. The MPEG-7 camera can be used for the tracking of the user position represented by the MPEG-7 *Moving Region DS*.

Within this work, we focused on the application virtual display. For many surveillance scenarios, only the motion of the objects is of interest. Therefore, only the positions of the moving objects are extracted. For this purpose, we implemented an efficient change detector [6] on the DSP of the MPEG-7 camera (see Fig. 2). The positions of the moving objects, described by the MPEG-7 *Moving Region DS*, are then sent to the surveillance operator. This approach allows us to fulfill one of the most important requirement that will encourage the deployment of surveillance systems, namely *privacy*. In surveillance applications many persons feel uncomfortable to be filmed. At the surveillance operator side, the virtual objects (in Fig. 2 represented by the blobs) perfectly follow the movement of the persons.

4. CONCLUSIONS

We have developed an MPEG-7 camera which can be used for a number of potential applications such as surveillance and augmented reality. One of the main advantages of the MPEG-7 camera as compared to a conventional camera consists in delivering only the relevant information to the targeted applications. Furthermore, the use of the MPEG-7

(a) Real scene

(b) Detection of moving objects

(c) Virtual objects

Fig. 2. Virtual display.

standard for the representation of the content-based description allows a modular design of applications thanks to interoperability brought by this standard.

We are working on a number of extensions to the MPEG-7 camera. Among these, the development and implementation of algorithms allowing the use of MPEG-7 camera for a mixed reality application called *CyberGlasses*. *CyberGlasses* is referred to a wearable display system in the form of glasses, capable of displaying virtual and natural objects in a mixed reality environment.

5. ACKNOWLEDGEMENTS

We would like to thank Fabrice Moscheni and the FAST*COM* team for their valuable inputs and help in the development of the MPEG-7 camera. Special thanks go to Emrullah Durucan for providing us with an efficient change detection algorithm.

6. REFERENCES

[1] T. Kurashige, J. Shiokawa, H. Chiba, N. Yamamoto, T. Kitade, H. Tarumizu, H. Kami, and T. Imai, "Development of MPEG camera," in *Proc. IEEE International Symposium on Consumer Electronics*, 1997, pp. 218–221.

[2] H. Katata, K. Hibi, M. Ito, T. Aono, and H. Kusao, "MPEG-4 camera for use with internet," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 3, pp. 661 –666, August 1999.

[3] "Introduction to MPEG-7," in *ISO/IEC JTC1/SC29/WG11/N4032*, Singapore, March 2001.

[4] "Multimedia description schemes FCD," in *ISO/IEC JTC1/SC29/WG11/N3966*, Singapore, March 2001.

[5] "MPEG-7 visual FCD," in *ISO/IEC JTC1/SC29/WG11/N4062*, Singapore, March 2001.

[6] E. Durucan and T. Ebrahimi, "Change detection and background extraction by linear algebra," *To appear in The Proceedings of IEEE, Special Issue on Surveillance*, 2001.