

Interoperable Affective Context Collection Using MPEG-21

Chris Poppe(i), Michael Ransburg(ii), Saar De Zutter(i), Rik Van de Walle(i)

(i) Ghent University – IBBT – IMEC, Gaston Crommenlaan 8, bus 201, B-9050 Ledeberg-Ghent, Belgium
E-mail: {[chris.poppe](mailto:chris.poppe@ugent.be); [saar.dezutter](mailto:saar.dezutter@ugent.be); [rik.vandewalle](mailto:rik.vandewalle@ugent.be)}@ugent.be

(ii) Klagenfurt University – ITEC, Universitätsstrasse 65-67, A-9020 Klagenfurt, Austria
E-mail: michael.ransburg@itec.uni-klu.ac.at

Abstract

The vision of the MPEG-21 standard is to specify a multimedia framework for the transparent delivery and consumption of multimedia content across a wide range of networks and devices. The user plays an important role in this framework. Therefore the optimization of the user experience is of uttermost importance. Context needs to be effectively collected and utilized to optimize this experience and MPEG-21 provides the necessary tools to do so. However, the efficient combination of these tools in order to create a context collection framework, remains an issue which is not tackled by the MPEG-21 standard. This paper proposes a framework, which incorporates affective context collection by combining several parts of the standard, as such creating an interoperable context-aware system.

Keywords: MPEG-21, affective context, DIP, DIA

1 Introduction

These days, the growing number of users in the world of multimedia computing causes a considerable increase in consumers and providers of multimedia content and applications. Moreover, users are becoming more and more critical about the applications they use and the way they can interact with them. Therefore, optimizing the user experience has become an important success factor for any interactive application. To maximize the user experience, various kinds of context information, such as the content preferences of a user, need to be collected, interpreted and processed [1]. Context collection has become extremely important [2][3], but only its integration into a complete multimedia framework can optimize its usefulness. The creation of an entire multimedia framework for the delivery and consumption of multimedia content across a wide range of networks and devices used by different communities is the goal of MPEG's latest standardization effort, named MPEG-21 [4]. As this promising standard will cover many domains of the multimedia world, it is important to find out how context collection can be incorporated in the framework. MPEG-21 standardizes normative descriptors for describing context in order to enable interoperable usage of this information.

However, the actual usage of these descriptors to define a context collection framework is not within the scope of the standard.

In this paper, we address the topic of collecting and utilizing affective context, stating the users' mood, and incorporating context collection into the MPEG-21 multimedia framework by combining several parts of the standard. The suggested system enables developers to plug their context collection tools into the framework with minimal effort and further enables transparent and interoperable usage throughout different MPEG-21 architectures. State-of-the-art context collection systems provide interfaces or ad-hoc context formats to which applications should be adapted. Our approach allows to combine context collection and MPEG-21 applications without altering them.

Section 2 gives a short introduction on affective context collection. In section 3, we briefly elaborate on some parts of the MPEG-21 standard, which are relevant for the proposed system. The created system is discussed in section 4 and consequential results and advantages are shown in section 5. Finally, section 6 formulates some conclusions.

2 Affective context collection

A normal human-to-human interaction will usually not only consist of vocal communication, but will be affected by the subjects current emotional state. Although vocal communication can be a direct measure of this state, several other signals, related to these emotions, can be captured and interpreted. An application, based on human interaction, can be abstracted to the following concept: user input is processed to generate output. Recently, we notice a shift in this user input. More and more systems are developed to apply a more human approach in gathering the input. Contextual information related to the emotional state of a person is called affective context and can be used to make an estimation of the actual emotional state. The research interest in affective context and its collection by measuring the emotion of a user, is increasing rapidly [2][3].

One can see that current affective context systems have a more or less similar approach in their operating. Mechanical

sensors, preferably non-invasive, are used to capture different kinds of values, such as skin conductance and temperature. These represent physiological changes, which are linked to emotional responses. The produced analogue signals are coded and by using training sets and learning systems, an estimation of the user's mood is made. This context is then put in an appropriate format for further usage. An example is given in Figure 2, showing a collected context in XML format stating the user was happy at a specific date and time with a probability of 72 percent. Additional parameters and elements could be used, such as type and version of the sensors, extra information about the environment or system, identifying information about the device, user or session, etc.

```
<Mood xsi:type="MoodType">
  <Emotion accuracy="0.72">happy</Emotion>
  <Date>2006-02-08</Date>
  <Time>15:22</Time>
</Mood>
```

Figure 1: Estimated Mood Example

Existing affective context collection applications collect and represent information in a proprietary format, which makes integration into a larger architecture difficult. Since research in the field of affective context concentrates on the accurate extraction of human emotion out of physiological measurements, not much work has been done to see how to effectively use the data in an interoperable way. In this area, the combination of affective context collection with MPEG-21 can create substantial benefits, i.e. the interoperable usage of the affective context in order to optimize the user experience.

3 MPEG-21

The MPEG-21 standard has been split up in several parts, which can be used individually or combined. Our system is based on two parts, named Digital item Processing and Digital Item Adaptation.

3.1 Digital Item Processing

In a MPEG-21 framework, the fundamental transaction unit is a Digital Item (DI), which can contain multimedia content, metadata and ways to interact with the DI. The possible interactions with a DI are standardized in part 10 of the MPEG-21 standard, named Digital Item Processing (DIP) [5]. DIP is to Digital Items as JavaScript is to HTML. A content author will, upon creation of a DI, insert DIP functionality, more specifically methods composed of script code, in the file. These methods can consequently be executed on any DIP-compliant terminal. DIP defines specific multimedia functionality to increase the possibilities of the methods included in the DIs. An API for a basic set of functions has been standardized to be used by the methods. A DIP-compliant terminal should provide an implementation for these interfaces, so similar behavior on different terminals can be exploited. As a result, a content author has means to write functionality in script code, and can call, from inside this

method, basic functionality available on the terminals. If the script code is insufficient for specific tasks, DIP has provided ways to insert externally generated code, in a pre-defined programming language, in the DI.

Consider a content author creating a DI, which represents a CD collection. The DI will typically contain references to music files, textual descriptions, CD-covers, etc. The content author might add some DIP methods in the DI which will play the music files in a pre-defined order, whilst creating a slideshow of the covers. A basic function has been defined to play multimedia resources, so the method can make use of this to play the music file. If the content author has specialized software to make a slideshow from a sequence of pictures it can include this in the DI as externally generated code and consequently use it to represent the graphical elements.

3.2 Digital Item Adaptation

Digital Item Adaptation (DIA), part 7 of the MPEG-21 standard, specifies a generic media bitstream adaptation framework, which enables coding-format and device independence. The former is enabled by a Bitstream Syntax Description (BSD), which provides a high-level description of the bitstream structure, and a Steering Description, which conveys possible configuration settings for an adaptation engine, based on the given Usage Environment [6]. The Usage Environment Descriptions (UEDs) include the description of terminal capabilities, network characteristics, user characteristics, and characteristics of the natural environment. As such, it not only provides device independence (i.e., the ability to play content independently of the presentation device), but also provides a means for describing user characteristics and preferences, such as the preference for a specific genre of content. By providing standardized descriptions of the user and his environment, an MPEG-21 adaptation engine can adapt the content accordingly. Therefore, the UED contributes to maximizing the user's overall content consumption experience. To use affective context in an MPEG-21 framework, one needs to convert the collected proprietary context into standardized descriptors. Our system presents a method to turn affective context into normative UEDs describing the users' preferences. This information can be used to provide advanced services, such as the adaptation of a movie by improving the quality of interesting scenes.

4 Context collection framework

DIP was created for the processing of DIs to create advanced interactive multimedia systems. However, we propose to go beyond the main purpose of DIP and use it to implement the context collection, hence benefiting from the interoperability provided by this normative API. We use the DIP functionality to implement context collection tools directly in a DI. Moreover, by creating appropriate wrappers, existing context collection tools can be incorporated into DIs, without changing tools. As such, context collection is made executable on any DIP-compliant terminal receiving these DIs.

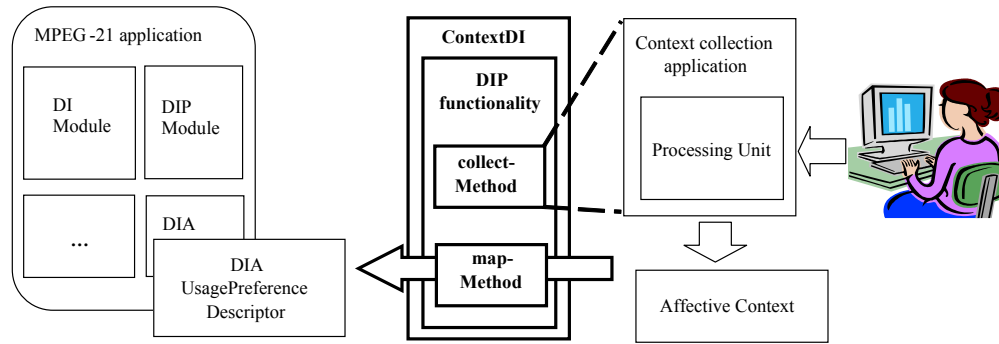


Figure 2: MPEG-21-based context collection framework

The steering of the context collection from inside the DI allows to configure parameters and collect context at any desired time.

Figure 2 shows the architecture we put forward to incorporate context collection in MPEG-21 and map the context onto DIA descriptors. In this scenario, we assume a user, interacting with an MPEG-21 application, who consumes some content. Affective context is collected by one or several context collection applications and is, after mapping to DIA descriptors, used throughout the MPEG-21 application. In our approach, a ContextDI acts as a middleware layer (represented by the bold parts in Figure 2) between MPEG-21 and context collection applications. We go beyond the core idea of DIP and use it to encapsulate the context collection into the DI (*collectMethod* in Figure 2). The execution of DIP methods, present in a DI (like the *collectMethod*), can be started by the user, the DIP application, or another DIP method. This way, the context collection can be executed at any given time (for example only when a user is watching a movie). The context collection application will typically generate proprietary output, representing the users' emotional state (as seen in Figure 1). Since the collection is incorporated in the DIP functionality, the output will also be available to DIP. We use another DIP method to transform this affective output into normative UEDs (*mapMethod* in Figure 1). Our system matches the affective data, representing the emotional state of the user, to a UsagePreferences UED. This is a description of the users' preferences and can be used to identify the preferences of a user for certain genres. Figure 3 shows an example of a UsagePreferences UED (namespaces have been disregarded for layout purposes). It expresses several preferred genres, namely sports, entertainment and movies. The figure shows that MPEG-7 descriptors [7] are used inside the DIA descriptor. The MPEG-7 user preferences tool facilitates the description of users' preferences pertaining to consumption of multimedia content. The DIA descriptor uses MPEG-7 constructs to describe the usage environment for adaptation purposes. Since the format of these descriptors is standardized, they can be understood and used throughout the entire MPEG-21 framework.

If a link has to be established between this description and the users' emotional state, information is needed on the actual content a user was consuming. In an MPEG-21 application, DIP will normally be the steering module for the processing

of multimedia content (for example, a movie DI will typically contain a DIP method, which starts the rendering of the movie, if requested by the user). Since DIP has access to the entire DIP application on which a DI is run, additional information can be collected about the context collection itself (for example the exact point in time of context collection).

```
<DIA>
  <Description xsi:type="UsageEnvironmentType">
    <UsageEnvironmentProperty xsi:type="UsersType">
      <User>
        <UserCharacteristic xsi:type="UsagePreferencesType">
          <UsagePreferences>
            <mpeg7:FilteringAndSearchPreferences>
              <mpeg7:ClassificationPreferences>
                <mpeg7:Genre>
                  <mpeg7:Name>Sports</mpeg7:Name>
                </Genre>
                <mpeg7:Genre>
                  <mpeg7:Name>Movies</mpeg7:Name>
                </mpeg7:Genre>
              </mpeg7:ClassificationPreferences>
            </mpeg7:FilteringAndSearchPreferences>
          </UsagePreferences>
        </UserCharacteristic>
      </User>
    </UsageEnvironmentProperty>
  </Description>
</DIA>
```

Figure 3: UsagePreferences example

Subsequently this data is used to determine the specific content or scene, which the user was consuming when his emotions were measured.

There is need for the explicit understanding of the semantics of a video(scene) if we want to match the semantics with a preferred genre of the user. To annotate video, we use MPEG-7 descriptors. As such, we can describe a specific (part of) a video and assign it to a certain genre, or even to multiple genres (for example, sports and movies). Thus, a link can be established between the description of the content and the UED descriptor. Our system introduces a method, which manifests this link by mapping the affective context onto UED descriptors. DIP offers functionality to process XML and eases the mapping. At this point we have an entire

sequence, starting with the collection of affective context and ending with normative UED descriptions.

5 Results

The contribution of this paper is the definition of a new architecture allowing to incorporate affective context collection into the MPEG-21 framework, by introducing a novel interaction between the different parts. By using a DI as a middleware layer, we create an interoperable affective context collection, framework in which (existing) context collection tools can be combined with MPEG-21 applications without having to alter either of them. DIP offers extended functionality available on any MPEG-21 DIP platform. Smart reuse of this functionality decreases the programming effort and increases interoperability of the system. Our system provides transparency on the actual context collection and is, not limited to affective context collection. Similar DIP methods can be created to include other kinds of context applications. However, to maintain interoperability, it is obligatory to map the collected context onto normative descriptions. The use of UsagePreference descriptors allows the introduction of personalized service delivery. DIA provides the necessary tools to automatically adapt content based on these descriptors. As such, it is possible to avoid scenes, which are not interesting to the user, or contrarily, show certain scenes at maximum quality. The knowledge about the users' emotions related to a certain movie(scene) can be exploited in numerous ways in any advanced multimedia architecture.

The architecture discussed here is a result of research done in the DANAE project (IST-1-507113; <http://danae.rd.francetelecom.com>). In this project the authors of this paper have contributed to the creation of a context-aware environment, allowing dynamic and distributed adaptation of scalable multimedia content. Various context is collected within the environment and consequently formatted, transported, stored and used to scale specific multimedia content. The project also showcases affective context collection, in which a platform has been developed, called AffectiveWare, which gathers data from physical sensors to make an estimation of a user's emotions. The topics of this project enforce the relevance of our system.

6 Conclusions

In this paper, we introduced an architecture to insert affective context collection in a transparent and generic manner into the MPEG-21 framework. The purpose of DIP is to allow content authors to specify the way their content should be processed. DIP specifies functionality available on DIP-compliant devices and standardizes ways to include interaction in the multimedia content. We utilize DIP for different purposes, by applying it to facilitate interoperable affective context collection. This reuse of DIP in a different context makes sense since context collection is a inherent part of a fully dynamic context-aware multimedia architecture. The system, described in this paper, uses DIP to encapsulate

affective context collection tools and to create standardized descriptions of the actual gathered context. The affective context is mapped on standardized DIA descriptors, describing the preferences of the user for a specific genre. The use of DIP allows to make the context collection executable on DIP-compliant terminals. By using this functionality, already available in MPEG-21-compliant systems, we have lowered the programming effort to a minimum, while staying away from ad-hoc solutions. The representation of the collected context in DIA UsagePreference descriptors allows for interoperability of the output from the context collection tools. It makes the affective context directly usable in an MPEG-21 environment and introduces advanced context-aware services which improve the user experience.

Acknowledgements

The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), Interuniversity MicroElectronics Center (IMEC), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO/Flanders), the Belgian Federal Science Policy Office (BFSP/O) and the European Union.

References

- [1] Dey A. K., "Understanding and using context", *Personal Ubiquitous Computing* 5(1), Springer-Verlag, London, UK, pp.4-7, February 2001.
- [2] Jonsson M., Harris H., "Emotions, Driving and In-car Speech Based Information Systems", *The Role of Emotion in Human-Computer Interaction*, September 2005.
- [3] Goulev P., "Utilizing real time AffectiveSensors to incorporate emotions into human computer interactions", *International Workshop on Wearable and Implantable, Body Sensor Networks*, London, 6-7 April 2004.
- [4] Burnett I., Pereira F., Van de Walle R., Koenen R., "The MPEG-21 Book", John Wiley & Sons Ltd., New Jersey, 2006.
- [5] De Keukelaere F., De Zutter S., Van de Walle R., "MPEG-21 Digital Item Processing", *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 427-434, June 2005.
- [6] Devillers S., Timmerer C., Heuer J., Hellwagner H., "Bitstream Syntax Description-Based Adaptation in Streaming and Constrained Environments", *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 463- 470, June 2005.
- [7] Manjunath B. S., Salembier P., Sikora T., "Introduction to MPEG-7: Multimedia Content Description Interface", John Wiley & Sons Ltd., New Jersey, April 2002.