User-centered multimedia content personalization: a services composition based approach

Marcelo G. Manzato
Mathematics and Computing Institute – University of Sao Paulo
Av. Trabalhador Sancarlense, 400, PO Box 668 – 13560-970, Sao Carlos, SP – Brazil
mmanzato@icmc.usp.br

ABSTRACT

In this project, we propose a distributed architecture which supports metadata extraction by exploring interaction mechanisms among users and content. The interaction activity addressed in this work is related to peer-level annotation, where any user acts as author, being able to enrich the content by making annotations, using, for instance, pen-based devices. The exploration of interaction mechanisms mentioned here will be implemented as services compositions, as well as the provision of adaptation tasks, so that users will be able to use portable devices to access and interact with multimedia content.

Categories and Subject Descriptors

J.9 [**Mobile Applications**]: Multimedia applications and multimedia signal processing; I.2.13 [**Artificial Intelligence**]: Knowledge Management—*Knowledge personalization and customization, knowledge acquisition, knowledge modeling*

General Terms

Design

Keywords

Personalization, adaptation, interaction, services composition, peer-level annotation.

1. INTRODUCTION

One of the greatest advantages that digital television has brought to users is the capability to interact with the content [15]. This affirmative can be supported if we look at the huge advances of the web, whose main characteristic is the interactive scenario where users are able to choose alternative navigation paths, explore different pieces of information and click on related links they wonder to check. The interaction with multimedia content, in special, is obtaining extra development efforts, because of the recent availability of web-based authoring tools, such as YouTube, Facebook, etc. Those tools have changed the user-consumer role into a user-producer role, enabling any user to create multimedia content and make it available on the web.

In parallel with the development of interactive web applications, it is becoming usual the use of portable and mobile devices to access multimedia content. This is due to the increasing processing power of those devices and because multimedia content can be manipulated in the sense of being adapted according to some particular situation. This last possibility, indeed, is issue of the content adaptation and personalization research area, whose adaptation is defined by Lum & Lau [12] as the applications' ability to choose the best version of the content to be successfully accessed by users with devices containing restricted capabilities. The personalization, in turn, is defined by Barrios et al. [4] as a particular case of multimedia adaptation where the data is adapted according to the needs and preferences of a specific user.

The multimedia adaptation and personalization, the use of portable devices, and the interaction functionalities have a relationship that is worth to consider. Firstly, users are more likely to interact with the content if they are using portable devices. Secondly, to make annotations onto multimedia content using portable devices, it is needed some kind of content adaptation, so that it can be successfully accessed and annotated by the users [12]. Lastly, as personalization services provide adapted content according to a user's preferences and needs, which may require a previous semantic knowledge about the environment in order to decide the best version of the content, the extraction of this high level information may be better accomplished if we explore some clues given by the user at the interaction time.

The interaction itself, as argued by some authors [7], can be classified in content consumption, setting selection, navigation/selection and authoring. This last category, as previously mentioned, obtains extra attention from users as it allows them to act as content providers and personalize the data, adding valuable metadata as additional information semantically related.

Usually, the authoring process can be done following two different approaches [5]: hierarchical, which provides information about specific media items with the objective to be searched or analyzed. One example could be metadata referring a movie, such as title, producer, list of actors, etc. The second authoring approach is called peer-level annotation, and can be accomplished by any person. One example mentioned by Cesar et al. [7] is the highlight of an actor's

text name using electronic ink. Two main characteristics of this level of annotation are that it does not follow a restrictive vocabulary [7], and that usually the annotation is made using portable devices, which have limited capabilities.

The need of semantic metadata to create personalization services certainly can be supported by hierarchical content authoring; however, the literature [1][18] reports that the job of annotating is time-consuming. Furthermore, producers will annotate certain characteristics that he/she subjectively thinks it is important for future applications.

Consequently, the exploration of peer-level annotation may minimize the hierarchical problems mentioned above. Firstly, the time-consuming question can be partially solved if multiple users dedicate collaborative effort to the same content. Secondly, the subjective choice of different aspects of the content to be annotated has a slighter drawback if we consider that the information extracted by the user interaction will be used for personalization services, which, in turn, will benefit the same user or set of buddies that usually shares the same preferences.

However, the use of peer-level annotation brings challenges that need further research. One of them is the fact that interaction between user and content can be done in different ways. Thus, a number of techniques must be available in order to analyze the interactive scenario and find valuable information that can be considered metadata. This set of techniques, indeed, sometimes needs to be combined with each other depending on the interaction activity executed by the user in a given period.

Another challenge is the provision of adaptation services, together with interaction functionalities, that must be considered when peer-level annotation is accomplished using portable devices. In this case, the complexity increases at the moment that the content to be accessed by the user has time constraints, such as interaction's response time or real-time applications.

Lastly, the peer-level has the characteristic of not following a restrictive vocabulary; consequently, algorithms to convert interaction-based extracted data into a representative format must be developed.

2. OBJECTIVES

As users usually enjoy to interact with the content, making personal annotations, using, for instance, pen-based devices, and maybe they wish to share those annotations with friends, this project proposes a distributed architecture with services compositions to support automatic metadata creation based on peer-level annotation at user/content interaction time. This information will, in turn, act as meaningful metadata for application scenarios, such as multimedia personalization. The main advantage of this approach is that users won't be bored as they usually do when using human-assisted tools to create content metadata, because they will be enriching the content without worrying about the true application's intention behind the interface. When necessary, the system will also be able to combine and use adaptation services to make possible the access of multimedia content using portable devices.

Considering the challenges of using peer-level annotation to extract metadata, the proposed architecture will address them as following:

- Combine different techniques (or services) that explore particular interaction aspects and adaptation procedures, such as handwriting recognitions, highlighted object segmentation, low-level features extraction, multimedia content transcoding and transmoding, etc.
- Isolate the unrestrictive annotation vocabulary by adding a representation model for metadata extracted from the interaction. This modeled information will then be available to any application that wishes to use them for personalization tasks.

3. RELATED WORK

The generation of metadata is a time-consuming task, and hence, should not be delegated only to content providers, as more content will be available each day. In addition, some authors report that users are usually loath to do manual annotation, and also, automatic analysis doesn't reach powerful results for the applications' goals which are wanted to have [19][6]. Consequently, the insertion of consumers into the metadata authoring process is a interesting design plan, being mentioned by many publications available on the literature.

Nack & Putz (2001) [17] and Gemmell et al. (2005) [11] use additional equipment and processing capabilities to capture conceptual dependencies at specific times of video creation. If on the one hand metadata can be extracted from these supporting technologies, on the other hand the user must be able to handle all the computational environment.

Considering this drawback about the need of user intervention, some pieces of work explore the fully automated metadata extraction activity. Lots of authors have published work that proposes techniques to gather related information about the content. Venkatesh et al. (2008) [19] cite some ways to extract this information; however, most techniques has good performance only in specific video domains.

As an attempt to solve the specific video domain problem, some authors [9] have published papers that deal with media aesthetic [20], which is a study and analysis of media elements such as lighting, motion, color, and sound both by themselves and their roles in synthesizing effective productions. However, each person will interpret the content information in a different way, and thus, users which don't have the same sensori-emotional values as those defined by the technique will have the possibility to not being satisfied.

Considering the fact previously mentioned that users may have different reactions for the same content, some work tries to model the user behavior in order to achieve semantic information extracted from multimedia content. Most work [10][8] models the user activity; however, they do not study the user annotation when interacting with the content.

Following research that deals with user models, authors have focused on user-centered approaches in order to provide personalization tasks according to personal preferences. Some publications, for instance, propose techniques based on personal traits [2]; but required mapping between user's preferences and audiovisual content features is not a trivial task to be accomplished.

Visual features, in special, are used in [3] in order to improve the metadata generation in a collaborative way. However, more study is necessary in order to determine when and how collaborative authoring and automatic metadata extraction can be all combined [19].

For collaborative work, user-produced data must be avail-

able for sharing in a way to be explored by applications. Content enrichment is explored in an architecture proposed by Cesar et al. (2006) [7] which supports authoring of additional information by users, and sharing of this data among users; however, no metadata is created by exploring the additional content which is inserted by the user.

At this point, considering the state of art previously depicted, it is possible to note that the use of peer-level annotations is suitable to extract meaningful metadata from multimedia content. The main techniques' limitations, which were described in this section, and that can be further explored are: i) the lack of techniques that use collaborative work from different people to annotate content; ii) the time-consuming efforts to author metadata; iii) the content-centered approach which generalizes user's perceptions at visualization time; and iv) the domain-specific techniques which are supposed to be fully automatic.

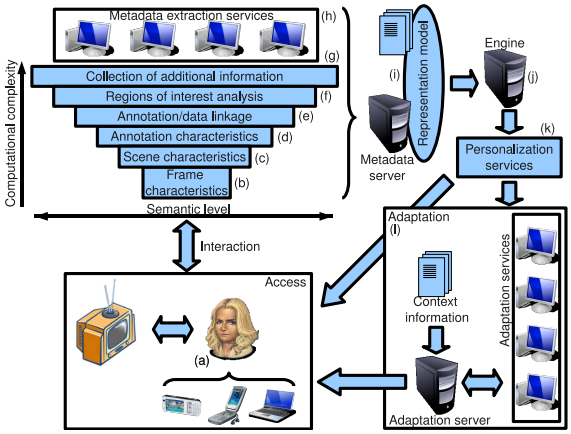

Figure 1: Proposed architecture for peer-level-based personalization.

4. METHODOLOGY

The methodology, in this text, is defined as the description of how the project could be developed. Firstly, this section depicts a set of techniques that can be used to extract metadata by exploring peer-level annotation. Lastly, an initial version of the architecture is presented, which can contribute to better understand the purposes of the work, and how they will be achieved. It is important to mention that partial results of the ideas presented here are already available on a publication [13].

4.1 Metadata Extraction

In the following subsections, we selected a set of issues that can be used to gather additional information from the user as he/she is interacting with the content.

4.1.1 *Individual frame(s) or scene(s) chosen by the user to annotate*

Low-level characteristics of chosen frames/scenes gathered from a video sequence may give information about content that is interesting to the user. These low-level characteristics may include histograms and color information, motion-field among successive frames, presence or absence of specific objects, etc.

Furthermore, users may restrict the access of specific scenes, once it may contain, for instance, inappropriate content for their children, or it may contain subjects that are out of scope of their preferences. Using classification techniques [14] [16], it is possible to define what kind of content will probably be denied or accepted by the user.

4.1.2 *Annotation gestures and assistive segmentation*

Related work about object segmentation usually has drawbacks, such as over-segmentation, when using those techniques to extract objects of interest without user's assistance. As users normally make annotations near interesting objects, some algorithm may be developed to assist the segmentation procedure in the sense of finding a relationship between the wanted object and the annotations nearby. As soon as the interesting object is extracted, the personalization task, in turn, may be beneficed by object recognition and feature extraction algorithms.

When considering only the annotation activity itself, some handwritings' characteristics, such as shape, intensity and speed, may give information about the user's sensation, feeling and level of interest about the content. For instance, circumferences around a specific area may imply that an important piece of information is present into that circle.

4.1.3 *Content sharing among users, documents and interfaces*

Usually, users have friends with similar tastes. Shared information gathered from a buddy may be used as additional medatada to identify the user's preferences. Discovering this relationship by comparing related documents may also inform what subject has triggered the user's attention, and hence, personalization tasks may explore this data in order to provide content according to user's tastes.

In addition, together with handwriting recognition, a number of different interfaces may be used for interaction. Voice recognizers, mobile phones, remote controls, tablet augmenters, etc., all of them have specific characteristics that can be analyzed in order to extract information about the user's preferences.

4.2 Initial Architecture

According to the different peer-level-based metadata extraction methods that were depicted in last subsection, we describe in this subsection an architecture that gives support to the exploration of interaction capabilities among user and content. Figure 1 presents the overall schema. Users may use personal devices, together with the standard digital televisor device in order to watch the presentation, at the same time that he/she is able to capture a frame, and make annotations using, for instance, a pen-based device (Figure 1 (a)). The interaction activity is monitored by a set of techniques that stores locally specific data, such as characteristics of frames (histograms, color level, etc.), captured frames IDs and timestamps from scenes, coordinates of strokes, recognized handwritings, etc. All this information is classified into different layers, represented by the blocks composing the inverted pyramid in Figure 1 (b)-(g), which, in turn, represents the required knowledge about the content.

The whole set of techniques corresponds to a services composition schema (Figure 1 (h)), which is controlled by a metadata server, responsible to manage, collect and store in a representation model valuable information about the

interaction activity among user and content (Figure 1 (i)). One example of such composition is the combination of segmented scenes with the annotations made onto a specific frame; this association, specially, is done by the comparison of the timestamps present in all media.

After all metadata is available in the representation model, the engine module (Figure 1 (j)) processes the data, creating the means for personalization services (Figure 1 (k)). Therefore, when the user interacts with the content, the architecture is able to extract metadata from specific information that is generated by the techniques. The metadata, in turn, is used to inform content providers about user's preferences that can be later explored to support personalization services.

The multimedia content that is transmitted from the content providers, which can be in its original form or personalized, is also processed by an adaptation procedure (Figure 1 (l)). Firstly, this step considers the collection of context information from the environment, and later, the multimedia stream is adapted using a set of composite services managed by an adaptation server.

5. EXPECTED RESULTS

Among the expected results of this project, we can highlight the following:

1. Implementation of a prototype that adapts and personalizes multimedia content according to environment's restrictions and user's preferences.
2. Interaction capabilities between adapted/personalized content and users, using different devices.
3. Possibility to extend the system by adding new adaptation and/or personalization services.

6. ACKNOWLEDGMENTS

This work is being sponsored by UOL (www.uol.com.br), through its UOL Bolsa Pesquisa program, process number 20080129100700. The authors would like to thank the financial support from UOL and FAPESP.

7. REFERENCES

[1] G. D. Abowd, M. Gauger, and A. Lachenmann. The Family Video Archive: an annotation and browsing environment for home movies. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 1–8, 2003.

[2] L. Agnihotri, J. Kender, N. Dimitrova, and J. Zimmerman. Framework for personalized multimedia summarization. In *Proceedings of 7th. ACM SIGMM Int. Workshop Multimedia Information Retrieval*, pages 81–88, 2005.

[3] M. Aurnhammer, L. Steels, and P. Hanappe. Integrating collaborative tagging and emergent semantics for image retrieval. In *Proceedings of WWWW 2006 Collaborative Web Tagging Workshop (Online)*, 2006. Available at: http://www.rawsugar.com/www2006/17.pdf.

[4] V. M. G. Barrios, F. Modritscher, and C. Gutl. Personalization versus Adaptation? A User-centred Model Approach and its Application. In *Proceedings of I-KNOW'05*, pages 120–127, 2005.

[5] D. C. A. Bulterman. Animating Peer-Level Annotations Within Web-Based Multimedia. In *7th Eurographics Workshop on Multimedia*, pages 49–57, 2004.

[6] D. C. A. Bulterman. Is it time for a moratorium on metadata? *IEEE Multimedia*, 11(4):10–17, 2004.

[7] P. Cesar, D. C. A. Bulterman, and A. J. Jansen. An Architecture for End-User TV Content Enrichment. *Journal of Virtual Reality and Broadcasting*, 3(9), 2006.

[8] T. Choudhury and S. Basu. Modeling conversational dynamics as a mixed-memory Markov process. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2004.

[9] C. Dorai and S. Venkatesh. Computational media aesthetics: Finding meaning beautiful. *IEEE Multimedia*, 8(4):10–12, 2001.

[10] N. Eagle. *Machine perception and learning of complex social systems*. PhD thesis, Massachusetts Inst. Technol., 2005.

[11] J. Gemmell, A. Aris, and R. Lueder. Telling stories with my lifebits. In *Proceedings of IEEE International Conference on Multimedia and Expo*, Amsterdam, The Netherlands, 2005.

[12] W. Y. Lum and F. C. M. Lau. A Context-Aware Decision Engine for Context Adaptation. *IEEE Pervasive Computing*, 1(3):41–49, 2002.

[13] M. G. Manzato, D. B. Coimbra, and R. Goularte. Multimedia content personalization based on peer-level annotation. In *Proceedings of the 7th. European Interactive TV Conference (EuroITV)*, 2009. (to appear).

[14] M. G. Manzato and R. Goularte. Video News Classification for Automatic Content Personalization: A Genetic Algorithm Based Approach. In *Proceedings of the XIV Brazilian Symposium on Multimedia and the Web (WebMedia)*, pages 1–10, 2008.

[15] M. G. Manzato, D. C. Junqueira, and R. Goularte. Interactive News Documents for Digital Television. In *Proceedings of the XIV Brazilian Symposium on Multimedia and the Web (WebMedia)*, pages 1–6, 2008.

[16] M. G. Manzato, A. A. Macedo, and R. Goularte. Evaluation of Video News Classification Techniques for Automatic Content Personalization. *International Journal on Advanced Media and Communications (IJAMC)*, 3(4), 2009.

[17] F. Nack and W. Putz. Designing annotation before it's needed. In *Proceedings of 9th. ACM International Conference on Multimedia*, pages 251–260, 2001.

[18] S. N. Patel and G. D. Abowd. The ContextCam: Automated Point of Capture Video Annotation. *UbiComp 2004: Ubiquitous Computing, Lecture Notes in Computer Science*, 3205:301–318, 2004.

[19] S. Venkatesh, B. Adams, D. Phung, C. Dorai, R. G. Farrell, L. Agnihotri, and N. Dimitrova. "You Tube and I Find" – Personalizing Multimedia Content Access. *Proceedings of the IEEE*, 96(4):697–711, 2008.

[20] H. Zettl. *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth, London, UK, 1999.