Multi-Sensor Information Retrieval for Lifestyle Applications

Hans van Gageldonk
Philips Research Laboratories
HTC-34, 5656 AE Eindhoven
The Netherlands
+31-40-27 47969

Hans.van.gageldonk@philips.com

ABSTRACT
Information and content is becoming abundantly available at any time, at any place. On the other hand, it becomes a burden for the end-user to find, organize, and enjoy content.

At the same time, people demand more personalized applications, given the strong societal trends towards individualization and people having ever busier lives.

This paper presents examples in content management applications for consumer devices that alleviate the search and retrieval problem for end-consumers. Then it is explained why body signal processing can be used to create truly personalized applications, where the end-consumer is fully immersed into the AV experience, while also offering means for active relaxation.

Categories and Subject Descriptors
J.7 [**Computers in other systems**]: Consumer products. H3.3 [**Information Storage and Retrieval**]: content analysis and Indexing – *abstract methods, indexing* methods. H5.2 [**information interfaces and presentation**]: User interfaces – evaluation, user-centered design. I.4.8 [**Image Processing and Computer Vision**]: Scene analysis - *color*

General Terms
Algorithms, Measurement, Performance, Reliability, Experimentation, Human Factors.

Keywords
Consumer electronics, content analysis, body signal processing, user study, video summarization, emotions.

1. INTRODUCTION
Information and content is becoming abundantly available. Audio, video, and pictures are instantly accessible and can be enjoyed on a multitude of connected devices. At the same time, it becomes harder for people to organize and retrieve the desired content at the right time and place. On the other hand, there is a strong trend in society towards individualization which raises an increased need for personalization. In addition, people live ever busier lives, and wish to have relaxed quality time with their family and friends.

In the Lifestyle program at Philips Research we explore opportunities in this space between information and content on the one hand, and truly personalized applications on the other. In this short paper I will highlight some examples of content management examples we have developed in the past. Further, I will motivate the transition to including body-signal processing as a new source of information, indispensible to create truly personalized applications.

2. CONTENT MANAGEMENT IN CONSUMER APPLICATIONS
Until recently, traditional consumer devices had no connection to the internet, except through the PC. This implied that content management applications could not be implemented in arbitrary devices in the network but instead had to be implemented on the device itself, e.g. on TVs, hard-disk recorders or portable media players. Such devices are heavily constrained in terms of computational resources, for they have to be price-competitive in the market. Typical hardware platforms consist of a dedicated processor and a limited amount of memory. Content management applications had to be designed such that they would “fit” the existing platforms: requiring a more advanced processor or more memory would immediately translate to a larger market price or a smaller margin. In addition, these applications were – and still are – subject to strong requirements in terms of being easy-to-use and robust. All these constraints make the design of content management applications on consumer devices very challenging. On the other hand, they also result in novel and smart implementations. An overview of some content management applications on consumer electronics devices has been published in [3].

Today we see a strong trend towards internet-connected consumer devices, for example in TVs. This opens the possibility of deploying content management and other applications (like games, e-mail, photo sharing, etc.) in virtually any device which is part of the network. On the other hand, the connected device will have to deal with content coming from all kinds of sources: broadcast via cable and satellite, internet content, photos from PCs, etc. Devices like set-top boxes, the SlingBox [1], and AppleTV [2] make first steps into this direction. Major companies active in the consumer electronics area, like Sony, Panasonic, and Philips, today bring products to the market with such functionality.

3. EXAMPLE APPLICATIONS
One of the early content management applications was the automatic detection of commercial blocks in a stored broadcast

stream. It uses automatically extracted audio and video features to determine where a commercial starts and where it ends. This allows the user to choose to watch commercials or not. This application has been implemented for example in hard-disk recorders.

A topic widely addressed in content management applications has been the automatic creation of summaries of video content. A typical application is the creation of a summary of a movie, as presented and discussed in [4]. This technology uses video and audio analysis to derive characteristics of each scene in a movie. Based on extracted characteristics, a summary of e.g. 1-2 minutes is created. An envisioned application is the usage in hard-disk video recorders, which can contain hundreds of hours of recorded video. When a consumer wishes to watch one of the many recorded movies, an automatically generated summary will offer him a way to get a flavor of the movie without giving away the full story. After watching the summaries, the consumer will then have a better overview of the available content and can more easily select a movie to watch. The main scientific challenge is how to create a summary that while being representative of the movie, does not give away the story. Many algorithms have been developed and tested by end-users to create a solid scientific base for this work.

Due to the often overwhelming availability of sports content, people often record sports broadcasts in order to watch them later. However, since the event already took place and sometimes they even know the result, they are typically interested in watching only the highlights, i.e., the most important parts of those matches. A system for automatically creating such summaries is described in [6]. Using audio and video cues it is possible to detect highlights in specific sports recordings. For example, in football matches it is possible to detect goals, missed chances, red and yellow cards, and other interesting events. On the other hand, algorithms that automatically detect such events cannot guarantee to deliver 100% correct results. Therefore it is essential to add a user interaction component to the system that compensates for this inherent limitation. This component can for an example allow users to watch highlights with various levels of detail of highlights (e.g. 10 highlights at level 1, 25 at level 2, and 50 at level 3).

To research whether people appreciate this system, and how they will use the interaction mechanisms, extensive end-user testing is important. This has been reported for soccer highlights in [7]. Figure 1 shows a screenshot of this application on a TV.

Figure 1: Example content analysis application: soccer highlight detection [7]

Automatic summaries can also be made of home video. Typically, people create many hours of home movies without bothering to edit the interesting scenes into proper video clips. Automatic video summarization can alleviate this task. A system called "Edit-While-Watching" has been implemented to allow the end-user to make small but important changes to an automatically generated home video summary [5]. This work also has been extensively tested by users.

There are many other interesting applications for content analysis in consumer applications, for example automatic mash-up creation of multiple video recording of the same event, e.g. weddings, music concerts [8].

4. TOOLS FOR USER-CENTERED RESEARCH

As explained in the previous section, extensive user testing is essential for a.o. content management applications. Automatic AV analysis algorithms cannot guarantee to deliver 100% correct results, and often it is not even clear what "correct" is. For example, people typically have different views on what actually a highlight in a soccer match is.

Philips Research has established a facility called ExperienceLab to do such user-centered research. ExperienceLab contains a home area, a shop area, and a hospitality area. The home area is a complete house, including a living room, a kitchen, and a bedroom, where people can try out new technologies and give feedback about usability, features, etc. It also contains means to observe people when they interact with the technology and applications. ExperienceLab allows researchers to get valuable feedback on the usability of new applications.

Recently we have created an on-line area in ExperienceLab, called SimplityLabs [9]. In SimplicityLabs we show software applications and concepts that people can use, after which we ask feedback on usability, interest, and other topics. Currently, the "Edit-While-Watching" application of automatic home video summarization and editing described in the previous section is available from SimplicityLabs.

5. BODY-SIGNAL PROCESSING FOR TRULY PERSONALIZED APPLICATIONS

As mentioned in the introduction, there is a strong societal trend towards individualization. Another trend is that people have ever busier lives and need means to relax themselves at home and also on the move. Complementary to that, there is also a trend towards the desire for more intense experiences, where immersion can play an important role; people would like to feel much more involved in e.g. movie-watching experiences.

This results in a strong trend towards personalization in consumer applications. Applications will need to be tailored to the person much more than today. One important aspect of this personalization is related to the current "emotional state" of a user. Understanding the emotional state can give a strong indication whether, for example, the person is relaxed or excited. Interpreting this information in real-time, can lead to radically new applications. An example of such an application is to adapt the amount of immersion effects (sound, video, light) of a movie to a person's mental and physical state: an intense, nerve-wrecking setting during a horror movie might not be the most adequate when the user feels mentally tired. Another example involves the creation of atmospheres in the living room by means of light. With information about the person's state of mind, it becomes possible to adapt the atmosphere to the user's desires and to better fit his needs.

Measuring and understanding a person's mental (and even physical) state is essential to enable these applications. The field of "affective computing" aims at such true personalization [10]. In general, emotions can be measured in two different ways: by detecting the facial expression of a user by means of cameras, and by measuring physiological signals [11]. Using cameras has the advantage that no body contact is needed and therefore is potentially more comfortable (albeit intrusive), yet it imposes restrictions on e.g. the position of the person in front of the camera. Measuring physiological signals, for example skin conductance response (SCR), heart rate variability (HRV) and electro-encephalography (EEG, brain signals) has been extensively used in laboratory settings and can give valuable cues regarding the user's state, but the step to use it in lifestyle applications has not yet been made.

A strong requirement is that measuring such signals needs to be as unobtrusive and comfortable as possible. Therefore the sensors should be placed in devices people will wear anyway, for example in a wristband of a watch (where one can measure SCR), a belt (measuring HRV) or in a headphone (EEG). Latest developments in sensor technologies allow for such integration. A big challenge is to develop new signal processing techniques to determine a mental state, for example whether a person is relaxed or not. These signal processing algorithms need to be robust and real-time. Latest developments in this area look very promising, but while the field for healthcare applications has met significant advances, the field for lifestyle applications is still in its infancy.

In the end, we envision systems that unobtrusively measure body signals, interpret those signals to determine a mental state, and feed that information in real-time back into the system. In such a way the system can for example use this information to transparently tailor the solution to real user needs and desires or even to give feedback to the user about his/her mental state, in a playful, enjoyable, and useful manner.

An example of such a feedback application is guided breathing; a system that measures heart-rate variability as a measure for relaxation. An example product in this area is the HeartMath personal stress eraser [12]. Feedback to the user about his current relaxedness state – measured in terms of heart-rate variability – can be given by means of light, but also in media experiences, for example by offering content that matches the current state of the person. The user can then actively relax himself by following the suggested breathing rhythm.

We envision a whole new area of applications when unobtrusive body signal processing in lifestyle settings can be done. Combining it with the knowledge that can be extracted from media, it allows for the step towards truly personalized applications. It allows people to be immersed into an AV watching experience tuned towards their needs and current state of well-being. In addition, people can also actively relax themselves in a playful manner, while enjoying their media experience.

6. CONCLUDING REMARKS

Today content is abundantly available. Content analysis and information extraction from content is an important topic for consumer applications to alleviate the task of searching and retrieving the right content at the right time. The next step is to create truly personalized applications by involving the current user's state of well-being into the media experience. Advances in sensor and signal processing technologies are very promising and will allow for such new lifestyle applications in the near future.

7. ACKNOWLEDGEMENTS

Mauro Barbieri, Pedro Fonseca, Marc Peters, Lu Wang, Bas Zoetekouw, Joyce Westerink, and Mark Johnson are gratefully acknowledged for their contributions to the work presented in this paper.

8. REFERENCES

[1] Sling Media. http://www.slingmedia.com

[2] AppleTV. http://www.apple.com/nl/appletv/

[3] Mauro Barbieri, Pedro Fonseca, Marc A. Peters, Lu Wang, *Multimedia Content Analysis for Consumer Electronics*, International Conference on Content-Based Image and Video Retrieval 2008 (CIVR2008), July 2008.

[4] Mauro Barbieri, *Automatic Summarization of Narrative Video*, PhD Thesis, Eindhoven University of Technology, November 2007.

[5] Campanella M., Barbieri M., Weda H., Kohlrausch A., Hoonhout J., *Easy Home Video Editing Following a User-Centered Approach"*, submitted to ACM Transactions on Computer-Human Interaction (TOCHI) Journal, January 2009.

[6] Bas Zoetekouw, Pedro Fonseca, Ralf Funken, *A fast and efficient highlight detector for soccer*, International Conference on Consumer Electronics, 2008 (ICCE 2008), January 2008

[7] Lu Wang, Pedro Fonseca, Bas Zoetekouw, *User Test of Soccer Highlights Application*, Proceedings of HCI 2008, September 2008

[8] Shrestha P., Barbieri M., Weda H., Sekulovski D., *Synchronization Techniques for Multiple-Camera Videos Using Audio-Visual Features*, submitted to IEEE Transactions on Multimedia journal, December 2008

[9] SimplicityLabs: http://www.simplicitylabs.net/

[10] R.W. Picard, *Affective Computing*, Boston MA, USA, MIT Press, 1997.

[11] Westerink et.al., *Emotion Measurement Platform for Daily Life Situations*, submitted to the 2009 International Conference on Affective Computing & Intelligent Interaction, Amsterdam, September 2009.

[12] HeartMath personal stress eraser. http://www.heartmath.com