Multimed Tools Appl
DOI 10.1007/s11042-011-0846-6

Discrimination of media moments and media intervals: sticker-based watch-and-comment annotation

Cesar A. C. Teixeira · Erick L. Melo · Giliard B. Freitas · Celso A. S. Santos · Maria da Graça C. Pimentel

Abstract In this paper we discuss the problem of how to discriminate moments of interest on videos or live broadcast shows. The primary contribution is a system which allows users to personalize their programs with previously created *media stickers*—pieces of content that may be temporarily attached to the original video. We present the system's architecture and implementation, which offer users operators to transparently annotate videos while watching them. We offered a soccer fan the opportunity to add stickers to the video while watching a live match: the user reported both enjoying and being comfortable using the stickers during the match—relevant results even though the experience was not fully representative.

This work has been supported in Brazil by grants from CAPES, CNPq, FAPESP, FINEP and MCT.

C. A. C. Teixeira · E. L. Melo · G. B. Freitas
Departamento de Computação, Universidade Federal de São Carlos, São Carlos, Brazil

C. A. C. Teixeira
e-mail: cesar@dc.ufscar.br

E. L. Melo
e-mail: erick_melo@dc.ufscar.br

G. B. Freitas
e-mail: giliard_freitas@dc.ufscar.br

C. A. S. Santos
Departamento de Ciências de Computação, Universidade Federal da Bahia, Salvador, Bahia, Brazil
e-mail: saibel@dcc.ufba.br

M. G. C. Pimentel (✉)
Departamento de Ciências de Computação, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, Brazil
e-mail: mgp@icmc.usp.br

Published online: 30 July 2011

Keywords Interactive multimedia · Interactive video · Digital TV · Sticker · TV · NCL · Authoring · Annotation · Watch-and-Comment · Interactive TV

1 Introduction

It is important to provide mechanisms for discriminating moments of interest in videos and live broadcast shows. For example, one may need to point to a particular moment in a video from a web-based video sharing system when delivering a lecture, or when creating a textual annotation or caption. As for live broadcast, providers may want to identify a particular moment in the video for advertising purposes.

There are many multimedia standards that allow the specification of moments and intervals in continuous media [1, 6, 54], methods for analyzing video content to identify events and activities of interest [12, 53, 57, 58], mechanisms to adapt video streams to different devices [19, 34, 57] and alternatives for the watching experience to be personalized by viewers [8, 11, 21, 22, 33, 36] or providers [3, 4, 15]. In the broadcast industry, it is a common practice to add extra content including animations and audio effects to live programs such as sports matches, and to recorded programs such as music videos. Similar approaches have been adopted by commercial video sharing systems. However, to the best of our knowledge, operations such as these where viewers can personalize their experience, or mark their moments of interest, have not been reported.

The *Watch-and-Comment* authoring paradigm has been presented as the seamless capture of a user's comments while watching a video and the association of such comments with the original media so as to generate interactive videos automatically, i.e. the original video annotated with the captured comments [7]. In previous work we defined *Watch-and-Comment* operators for the discrimination of media moments and of media intervals within continuous media, and showed how these operators can be used by applications processing annotations generated collectively by several users, over distinct instances of media and at different times [52].

The primary contribution of this paper is a system which allows users to personalize their programs with previously created *media stickers*—pieces of content that may be (whether temporarily or not) attached to the original video. We present the system's architecture and implementation, which offer users operators to transparently annotate videos while watching them.

To experiment with our approach, we built an application that allows viewers to add stickers to interactive video programs running on the Brazilian Interactive TV Middleware [49, 50], extended to support the discrimination of segments of media by multiple devices communicating via the UPnP protocol.

In a field trial, we offered a soccer fan the opportunity to add stickers to the video while watching his favorite team play an important match live on TV. After the match—during which the system worked as expected—the user reported having enjoyed applying stickers to personalize his TV watching experience, and offered several suggestions regarding both the stickers and the system: these are relevant results even though the field trial was not fully representative.

Our current proof-of-concept application uses only one predefined sticker: this should be overcome by a new application, which will allow the configuration of a set of stickers to be used during the broadcast of live shows. Moreover, in our current

prototype the discriminated moments and intervals of interest cannot be saved: this will be solved by the integration of the current implementation with services detailed in previous work [52]. Finally, future qualitative evaluations will consider carrying out the experiment with groups of users.

This paper is structured as follows: Section 2 presents a scenario employing stickers; Section 3 introduces the discrimination of moments and of media intervals with stickers; Section 4 discusses the prototype infrastructure we built to support the use of stickers, along with a proof-of-concept application; Section 5 details a field trial carried out with a real user; Section 6 reviews related works; and Section 7 presents our final remarks.

2 Motivating scenario

The scenario consists of a group of friends who gather to watch an important football match broadcast live, and for free, in the Brazilian terrestrial TV system. All friends are fans of Sport Club Bahia, and about to watch their team play against its biggest rival: Sport Club Vitoria.

2.1 Sports watching scenario

Dr. Prazeres, a lifetime fan of Sport Club Bahia, invites his friends to watch the championship's final match against chief rival Sport Club Vitória. He has bought a 52-inch TV to enjoy the Brazilian terrestrial digital TV system, which is provided with high definition content with customizable interaction services—all free of charge.

To prepare for the event, Dr. Prazeres buys not only food and drinks but also glasses with his team's logo. He hangs a huge flag of his team on the wall. As all of Dr. Prazeres's many friends arrive, he concludes the preparation of his TV set-top-box by downloading pictures, animations, and chant *stickers* published in the club website so as to used them during this particular match. Since several friends have smartphones, Dr. Prazeres explains how they can connect to the TV set-top-box and add stickers themselves.

The next couple of hours go as happily as planned. When his team enters the field, Dr. Prazeres proudly adds a sticker with the official anthem. During the match, each of the goals scored by his team is celebrated with stickers. During the half-time interval the friends agree that the stickers were intrusive in some moments: Dr. Prazeres brings a 26-inch TV set from his bedroom and the smartphones are set to present stickers on the smaller TV. At the end of the match, the happy fans learn that they can use the information stored in their devices—the moments and intervals in which they used stickers—if watching the same match again using a recorded version.

2.2 Discussion

The scenario illustrates a particular use of the *Watch-and-Comment paradigm* [7], and exploits the possibility of using a remote control or the users' own mobile device to interact with the video. In this case, the *comment* (from the watch-and-comment point-of-view) corresponds to the discrimination operations associated with the scenes.

It is important to observe that the discriminating information can be used at the time of capture, as in the scenario presented, or in future opportunities, when users review the best moments of the match, for example.

In both cases, because the selected moments may be recorded in the viewers' mobile devices, the applications can process the discriminated media moments and intervals in a distributed, synchronous or asynchronous way (see [52] for examples).

The adoption of personal portable devices such as smartphones demands the use of a protocol that allows the discovery and the negotiation of services automatically—as is the case with the UPnP protocol. It also demands the provision of accessible and intuitive user interfaces—which is always a challenge. Moreover, the overall communication and information capture and retrieval must be supported by the underlying middleware.

3 Discrimination of media moments and media intervals with *stickers*

Our interpretations for *continuous media moment* and *continuous media interval* [52] borrow concepts from Allen and Hayed [2], and are in line with standards that allow the specification of the temporal behavior of multimedia presentations.

3.1 Discrimination of media moments and media intervals

A *continuous media moment* is an *instantaneous* segment of the media which takes a point m_a as a media offset, and a brief tolerance Δ before and after the offset m_a, as indicated in Fig. 1(a). Similarly to Allen and Hayed's definition of *time moment*, a tolerance Δ defines the meaning of the word *instantaneous* in a given context—in our case, the context is continuous media. The discrimination of a *continuous media interval* may be achieved by the specification of a segment of media, a point m_b as the media offset, and a duration of $\beta - \alpha$, as shown in Fig. 1(b). Alternatively, it may be defined as a segment of media delimited by two time moments, m_c and m_d, as illustrated in Fig. 1(c).

The literature reports important results on how to represent moments and intervals. As far as declarative languages are concerned, both NCL [1] and SMIL [6] define operators to be used in the specification of the temporal composition of a multimedia document so that the presentation of each media element is associated with the presentation of the context in which it is embedded. As a matter of fact, according to Rogge et al. [47] SMIL allows authors to describe 29 multimedia interval relationships. As an example, in a SMIL presentation a `par` (or `seq` or

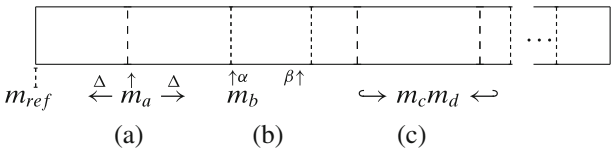

Fig. 1 **a** Continuous media moment given by media offset and tolerance Δ. **b** Continuous media interval with one media offset and duration of $\beta - \alpha$. **c** Continuous media interval with two media moments (each with its own defining media offset and tolerance)

excl) container can group, say, text, video or audio elements, and the duration of each element may be specified by the intrinsic duration of the media element, by the duration attribute (e.g., dur="5"), or by attributes giving the start and end moments of the interval (e.g., begin="4s" end="8s"). In the context of activation of a single media object, the clipBegin and clipEnd elements allow the specification of temporal subparts of a media object using the area element (e.g., <area id="subpart01" begin="4.1s" end="6.9s" />). These elements are also part of the proposal for the SMIL Timesheets specification [56], whose aim is to provide other XML languages with the timing control provided by SMIL.

Since NCL also has the power of expression to allow authors to specify the temporal behavior of more complex multimedia presentations [51] than the one demanded by our stickers, both NCL and SMIL offer alternatives for the specification of stickers as continuous media intervals within multimedia compositions. The examples discussed in this paper correspond to NCL documents which run in an NCL-based infrastructure [50].

The HTML 5 specification proposal, in progress [26], has elements to include video and audio as first class objects. Several proposals are under way to allow the actual use of moments and intervals so that captions and transcriptions, for instance, can be related to moments and intervals in video and audio files in an declarative format. One example is the track element, which is aimed at allowing authors to specify explicit external timed text tracks for media elements: the proposal of text track elements include text track cue elements which may contain start time and end time elements. Another example is the ongoing effort involving the specification of Media Fragments [54]. The work on Media Fragments investigates how to specify portions of a media object as a piece of media in one URI, which allows the play back of a specific portion of a video file from an external reference, or without reloading the page. The current proposal suggests the use of an abbreviated syntax that may or may not involve the associated Normal Play Time [31]—examples are t=10.20 for time moment, [10,20) and t=npt:120,0:02:01.5 for time interval [120,121.5). From the present work on HTML 5 and associated specifications, it is clear that there are efforts to provide specifications for time moments and intervals—what is not clear is whether the expressive power provided by SMIL and NCL will be supported. This is in line with the comparison provided by Laiola Guimarães et al. [33] on how SMIL, NCL and HTML 5 allow end-user time-based annotation.

3.2 Defining stickers with media moments and media intervals

In our scenario viewers can use media *stickers* to enrich their viewing experience, which means that multimedia transformations may happen during rendering time—a category of transformations which has been receiving increasing interest in the last years [32].

Stickers may have an intrinsic time duration (*interval-based stickers*) or have intrinsic time duration equals zero (discrete or *instant-based stickers*). Examples of *instant-based stickers* are graphics and photos; examples of *interval-based stickers* are animations and hymns. To be perceived by viewers, the presentation of a sticker must be associated with a time interval. Therefore, instant-based stickers must be

specified both by a *media moment*, as discriminated by users, and by a *duration*: the overall result in a continuous *media interval*.

For interval-based stickers, their presentation may be specified by a media moment discriminated by the user and by their intrinsic time duration. Moreover, the presentation of interval-based stickers may also be associated with an explicit duration, which may cause the playback of a sticker to be interrupted in case the specified duration is shorter than the original media's implicit duration.

In the NCL code presented in `Listing 1`, a sticker is specified by the combination of two media items: the Bahia's logo (`line 15: bahia.png`) and its famous anthem (`line 16: anthem-bahia.mp3`). The sticker specifies that both components should be presented in parallel (because all ports specified in the body (`lines 13-14`) are started at the same time), each playing for 15 seconds (`lines 7-8`). A similar document is used in the proof-of-concept application presented in Section 4.3, and in the field trial discussed in Section 5.

```
00
00 <!-- Listing 1: sticker with media/bahia.png + anthem-bahia.mp3
00
01 <ncl id="soccer" xmlns="http://www.ncl.org.br/NCL3.0/EDTVProfile">
02   <head>
03     <regionBase>
04       <region id="rgSticker" right="10px" top="10px"
                                width="150px" height="150px" zIndex="2"/>
05       <region id="rgAnthem"/>
06     </regionBase>
07     <descriptorBase>
08       <descriptor id="dSticker" region="rgSticker" explicitDur="15s"/>
09       <descriptor id="dAnthem" region="rgAnthem" explicitDur="15s"/>
10     </descriptorBase>
11   </head>
12     <body id="main">
13       <port id="pImg" component="img"/>
14       <port id="pAnthem" component="anthem"/>
15       <media id="img" src="media/bahia.png" descriptor="dSticker"/>
16       <media id="anthem" src="media/anthem-bahia.mp3" descriptor="dAnthem"/>
17     </body>
18 </ncl>
```

In order to make sticker presentations possible, the corresponding discrimination of media moments and intervals must consider the time relative to the presentation. In the context of digital TV, the timestamp needed may use the Normal Play Time (NPT) [31] associated with the media stream: this timestamp indicates the stream's absolute position relative to the beginning of the presentation. This solution, recommended to digital TV systems [48], is adopted in the Brazilian Interactive TV Middleware [50] and in the specification of the syntax for constructing media fragment URIs [54], to name a few.

4 Prototype implementation

To provide operators for the discrimination of moments and media intervals for various applications with different functionalities, we implemented: (a) a service which runs on top of the set-top-box, and (b) a service which runs on the user's mobile device. We observe that our current implementation reuses components that support the collaborative annotations of points of interest presented in previous work [44].

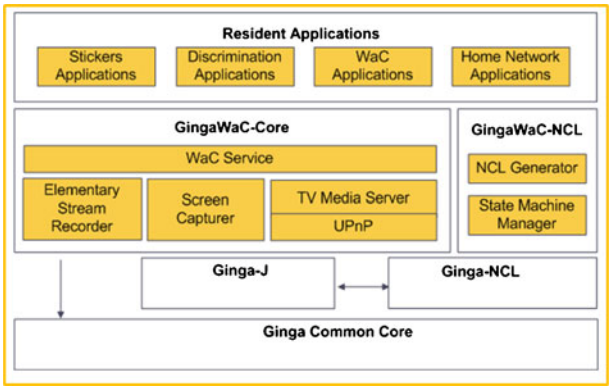

Fig. 2 Ginga-NCL extension: original Ginga modules are indicated in *white*, our modules are indicated in color

4.1 Discrimination service in the STB

A portion of our implementation runs on top of the Ginga-NCL middleware for digital TV [50], as depicted in the architecture shown in Fig. 2: the Ginga-NCL's original modules are presented in white, whereas our modules are shown in color. We have redesigned modules presented in previous work [52] so that our extension can be used in the Resident Applications Layer in a set-top-box with the Ginga middleware. Once installed, any application may take advantage of the UPnP protocol for communication with mobile devices, as well as of the Watch-and-Comment services supporting the capture of the live interaction between the user and the interactive TV applications, and of the presentation of associated interactive programs generated on the fly by the server's portion of application.

Our current implementation includes modules that work as follows:

- *Sticker applications* allow viewers to personalize their watching TV experience. These are in the same level as other applications that use the watch-and-comment paradigm such as the discrimination for later review [52] and services designed for the home [17];
- the *WaC Service module* offers the discrimination of moments and intervals in the network through the UPnP protocol, and allows resident applications to register themselves to be notified about user interactions;
- the *Elementary Stream Recorder module* provides for the recording of all elementary streams in a way that it can be referenced later for presentation;
- the *Screen Capture module* captures the selected video frame so that it can be presented later;
- the *TV Media Server module* provides audiovisual content transparently in the network;
- the *UPnP module* implements a code relative to the UPnP protocols so as to provide its functions to other devices throughout the network;
- the *NCL Generator module* produces an interactive multimedia document aggregating the information captured by the other modules;
- the *StateMachine Manager module* captures the current state of a presentation.

4.2 Discrimination service in mobile devices

Being able to run on mobile devices, the client portion of our prototype's implementation is a service that implements a *UPnP control point* responsible for locating and using the applications provided by TV set-top-box. The main advantage of exploring UPnP in this scenario is the use of the plug-and-play alternative to the communication network: this means that interactive applications can use the service without demanding that the users perform complex settings.

The overall architecture supports several types of devices, and the discrimination service running on the client device can be implemented in many programming languages. The proof-of-concept service we built was developed using the *Java Micro Edition* for mobile devices.

The main component in the discrimination service running in the client's device is the UPnP *control point*. It is responsible for locating a UPnP device on the network, and for allowing the viewer to use the services provided by those devices. When the discrimination service is initiated on the client device, the search for UPnP devices is automatically performed. If a device with a *Watch and Comment* service is found, the application automatically enables the viewer to use this service via a graphical interface on the device. As far as the use of UPnP as the underlying communication protocol is concerned, our work is mostly related to that reported by Holbling et al. [27], who reported multiuser access to interactive TV platforms using the UPnP architecture.

4.3 Working scenario

Considering sticker-based annotations like the ones provided by our prototype implementation, the scenario in which Dr. Prazeres and his friends watch the Bahia *vs.* Vitória match would work as follows:

- When Bahia scores its first goal, Dr. Prazeres uses the remote control to add stickers, as shown in Fig. 3a: an image containing the team's logo is added along with the Bahia anthem.
- Elsewhere, an upset Vitória fan adds a sticker with an ashamed version of his team's mascot, the lion (Fig. 3b).
- As the match continues, Vitoria's best player scores a goal. At this moment, Dr. Prazeres becomes so frustrated that he sticks an upset version of his team's mascot to the video while the opponent player celebrates his goal (Fig. 3c).
- Elsewhere, a Vitoria fan celebrates the goal by adding a sticker combining an image of his team's mascot and logo with the Vitoria anthem (Fig. 3d).

5 A field trial

In this section we report a field trial carried out with our infrastructure: a soccer fan who was going to watch an important soccer game was given the opportunity to add stickers to the presentation any time he wished. The viewer was a fan of Corinthians, who was playing against its chief rival, Flamengo.

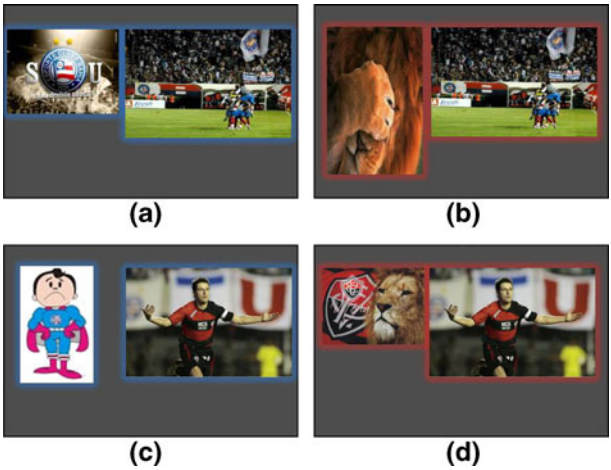

Fig. 3 *Stickers* selected by the users during the match: **a** when Bahia scores a goal, Bahia fans select a sticker supporting their team while **b** Vitória fans select a sticker which shows their sadness; when **c** Vitória scores a goal, Bahia fans select stickers corresponding to their frustration and **d** Vitória fans select stickers indicating their happiness

The fan watched the game in the comfort of his home. Along with our service, we had previously made a single interval-based sticker available in the TV set. This sticker, which was similar to the one presented in Listing 1, combined the club flag and the club anthem (see Appendix for the NCL programs). To add the sticker to the presentation, the viewer used the red button in the remote control. In other words, the activation of the button by the user caused the club flag to be placed on the screen and the club anthem to be played back—the club flag was presented on top of the video in parallel to the playback of the anthem, as illustrated in Fig. 4.

It is worth observing that, when preparing this setting, we decided not to use smartphones because we did not want its use to interfere with the treatment offered by the sticker. In other words, we wanted the sticker to be the only novelty for the user.

5.1 Observation

We observed that, at first, the user planned to activate the sticker only when a goal was scored—in his comments he was going to use it only after a goal by his team, and he expected his team to score lot of goals ("at least three" was his comment).

Fig. 4 *Stickers* added by a user while watching an important soccer match at home

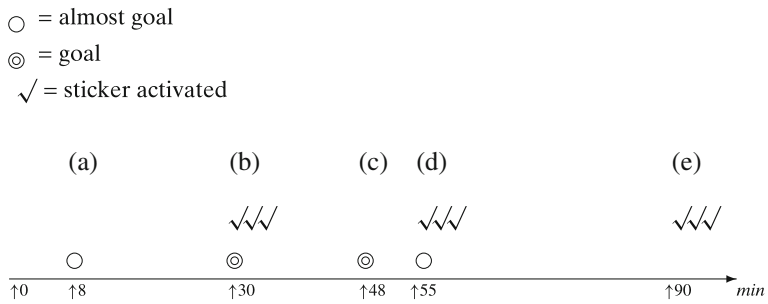

Fig. 5 Timeline illustrating important moments of the match. The symbol √ indicates time instants in which the fan activated the sticker

The first important moment in the match occurred when his team came close to scoring a goal, in the 8th min of the first half (Fig. 5a). The user did not use the sticker at this opportunity, as he had planned to use it only when there was a goal. When a goal was finally scored at 30 mins of the match, the fan celebrated and used the sticker. In fact, he activated the sticker several times after the first activation. It is interesting to note that, despite the fact there was only one sticker, the fan used it several times, which resulted in a longer intervention in the presentation of the match (Fig. 5b).

At 48 mins of the match, in the beginning of the second half-time, the fan became frustrated as the opponent team scored a goal (Fig. 5c). At 55 mins of the match, the fan enthusiastically used the sticker after his team had come close to scoring a second goal, had the ball not hit the pole (Fig. 5d). From that moment on, the fan activated the sticker several times. During the match he thought it would be good to have stickers to tease the other team in situations where it, or the referee, made mistakes. In the end, although no other goals were scored, the fan activated the stickers several times just for fun (Fig. 5e).

5.2 Questionnaire

After this experience, we asked the user to answer a questionnaire containing nine five-point Likert items and two open questions. When answering the nine items, the user informed that:

- Thinking about adding a sticker was not a concern during the match
- He was comfortable with adding stickers during the match
- The fact that the sticker added audio on top of the original audio was not a problem
- The fact that the sticker added an image on top of the video was not a problem
- We would not prefer using an anthem-only sticker
- We would not prefer using a logo-only sticker
- He would have used other stickers had they been available
- He added more stickers than previously expected
- He did not only add stickers in important moments of the match

In the first open question, the fan suggested other stickers he would like to use, such as: animated fireworks; images and audio stickers with complaints about the referee's decisions; and combined text and audio stickers to make fun of the other team.

In the second open question, the fan offered the following suggestions: (1) to provide animated stickers and stickers with transparency; (2) to reduce the original audio when stickers with audio are used; (3) to offer more stickers; and (4) to allow the customization of stickers.

5.3 Discussion

From his comments, we observed that the user enjoyed using the sticker. After applying it a couple of times, he decided to use it more often and suggested other stickers he would use had they been available.

When answering our questionnaire, the user confirmed our supposition that adding stickers during the match was not a problem, and that he was comfortable with doing it. However, although he said that adding the anthem on top of the original audio was not a problem, he suggested lowering the volume of the broadcast audio when audio-based stickers were used. Similarly, although he answered that adding an image on top of the original video was not a problem, he suggested using transparency when using image-based stickers. The user also made it clear that he preferred using combined stickers to single-media stickers and mentioned that it would be interesting to be able to personalize his own stickers rather than only having predefined ones. The fact that providers commonly use animated stickers on their programs may have inspired the user to request such animations to be made available to the end-user as well. Moreover, the request for personalization was expected as fans are used to personalizing their cell phones with ring tones from their teams[1] and the screen of their personal computers with professional wallpapers, for example.[2]

This experience shows a user enjoying adding stickers to a live program. In the field trial the system worked as expected, and the user reported being comfortable with using stickers during the match—even though the field trial is not fully representative, these are relevant results.

In many scenarios, it is important to register this user-interaction so that the discriminated moments and intervals are associated with the original media.

6 Related work

It is important to be able to identify which portions of videos or live broadcast shows are of interest to viewers. In the context of broadcast TV, for example, specialized audience measurement research institutes worldwide[3] must collect data regarding

[1] E.g. http://www.gsmfans.com.br/index.php?topic=602.0 currently offers ring tones from 39 Brazilian teams.

[2] E.g. http://www.ecvitoria.com.br/site/papeldeparede/default.jsp offers wallpapers of the Vitoria Sport Club.

[3] http://www.international-television.org/tv_audience_measurement_research_boards_and_institutes.html

which and where each of the many available channels is watched at any given time. If available, a list of moments of interest may be used by indexing, summarizing, annotation, recommendation or personalization services.

Personalization and collaboration has become a mandatory feature on several commercial video-related services. In the U.K., the BBC's Internet-based radio and TV on-demand platform, iPlayer,[4] has made efforts to provide users with a novel personalization interface, recommendation services integrated to social networks, and synchronous chat based on existing message exchanging services. Similar features are offered in the U.S. by the Hulu[5] video-on-demand service. In a different category, Boxee[6] is a software application (with optional hardware) designed for TV which enables connectivity to video servers for download and playback. As the other two, Boxee also allows users to view, rate and recommend content to friends via existing social network providers. Systems like these allow remote users to share their watching TV experience in a somewhat more structured way than they already do when watching TV and connecting on their favorite social networking and chatting services—the latter being investigated for the extraction of points of interest [55].

Regarding the identification of portions of interest, researchers investigate both preprocessing and live processing of videos, as well as the use of explicit or implicit viewer input. Preprocessing may involve the extraction and clustering of video shots to create a thumbnail version of the video [12], a video summary [24, 35, 40, 53], or a storyboard [23]. Preprocessing may also demand the tracking of recorded user preferences to provide recommendations [16, 20, 28]. Researchers also investigated the cropping [19] or the indexing and adaptation [34] of the video for various devices. The determination of events of interest that take implicit user input into account includes monitoring instant text [55] and multimodal messages [38] exchanged by viewers while watching the video, or monitoring viewers' facial expressions [37, 59]. The possibility of adding temporary stickers is a novel and ubiquitous form users have to indicate portions of interest in the programs they are watching.

Regarding personalization, there are several opportunities for users to personalize their watching TV experience—examples have been reported in the literature in the context of interactive television both on the server side [15, 30] and on the client side [13, 39, 45]. Early efforts were geared towards offering TV viewers extra related contents [21], or offering them the opportunity to skip TV content [22]. The selection of moments of interest also made it possible to investigate how to allow viewers to go back and forth on the video and to add annotations [36]. The need to offer possibilities of personalization has led to efforts that include offering powerful search mechanisms [41], integration with semantic web technologies [4], and the processing of metadata [30] to extract the emotional atmosphere in the media [42], among many other efforts involving mining in multimedia streams [58]. Advertising has also been a motivation for investigations involving personalization [3]. Personalized authoring by the end-user has also been investigated when content starts from scratch in mobile devices [5]. Personalization integrating interactive TV and the home environment

[4] http://www.bbc.co.uk/iplayer

[5] http://www.hulu.com

[6] http://www.boxee.tv

has demanded research involving mobile devices [17, 18, 27]. Considering the above, the possibility of adding temporary stickers is a novel form of personalization.

The association of additional contents to TV programs has been reported in the literature, as in the cases of the AmigoTV [13] and CollaboraTV [39] systems, but in theses cases the research focus was on television-based communication. Coppens et al. [13] allowed remote viewers using their AmigoTV system to add animations to their video—and the animation was shared among collaborating users. The presentation of extra contents associated with a program was also allowed in the CollaboraTV application [39]: the system allowed users to collaborate synchronously or asynchronously by selecting avatars to report the users' opinion by text or expressions associated with a particular time offset in the video. In both cases, the focus of the authors was on allowing communication among remote users. A recent work has proposed monitoring the communication among users in conventional messaging systems to allow the identification of events of interest within video content [55]. The possibility of adding temporary stickers is a novel form of associating additional contents to TV programs.

As far as the annotation on continuous media is concerned, the literature reports works with different purposes. Goularte et al. [25] investigate the use of annotation on video frames by pen-based electronic ink or voice recognition—the annotations are, in both cases, converted to text so that an XML document can be generated to reference all media corresponding to each annotation session. A comprehensive approach to enrich multimedia interactive documents has been proposed by César et al. with focus on secondary screens [8, 9]. In another research effort, Costa et al. [14] propose a model in which annotations are made based on temporal data and multi-tracks, and each track has a perspective of the same video content. In a multi-user environment, the collaborative aspect can be exploited through the tracks: users define tracks and then are able to make annotations. In a complementary approach to our work, Ramos and Balakrishnan [46] exploit various techniques of visualization and interaction for fluid navigation, segmentation, linking and annotating digital video by means of digitizer tablets supporting pen-based input.

In comparison with previous technologies, interactive digital TV has the potential to offer many types of user-video interactions and viewer-service interactions. Interactivity may occur, for instance, by means of access to external services, made possible by the computing power and communication capabilities that can be incorporated into TV sets. These features open new opportunities for research, since new types of services, applications and interaction paradigms can be exploited. Examples include the investigation of alternatives for supporting social interaction via TV-related asynchronous interaction among viewers by means of implicit recommendations of TV shows [39], as well as the use of synchronous explicit recommendation [13]. Another example is the study of communication choices and practices in TV-based text and voice chat [29]. Yet another example is the investigation of how passive users can interact more directly with the content they watch in TV-environments [10]. A more comprehensive study has involved four concepts for social sharing: fragment, tag, enrich, and send [11]. As a final example, our own previous work, the Watch-and-Comment authoring paradigm, proposes the capture of the comments made by viewers while watching TV programs, so that the comments are treated as multimodal annotations used to automatically create user-annotated interactive videos [7, 43].

A recent approach has extended the concept of collaboration with video even further: in the proposal by Jansen et al. [32], the interaction is centered around a distributed shared activity whose communication is based on video exchange. Supported by SMIL-based implementations, this work has led to a preliminary version of a new model for editing operations on active temporal multimedia documents.

Inspired by earlier results on the Watch-and-Comment authoring paradigm, in previous work we explored its approach in the definition of operators for the collective discrimination of moments and of media intervals within continuous media [52]—the aim was to allow the merging of annotations by individual users, for instance. We built applications in which users were expected to discriminate one scene when they recognized a particular value for the scene with respect to another scene, the value being associated with the context of the application. Compared with our previous work, as well as with other works related to annotation and to the identification of moment of interest, the main novelty of the approach presented in this paper is the ease-of-use to discriminate moments of interest during a program and the capture of the corresponding moments and intervals.

7 Final remarks

The sticker-based application described in this paper is an example of how users can engage with the possibility of interacting directly with the content they watch and comment in informal settings such as their homes. From a personalization perspective, our stickers are similar to the feature commonly offered by webcams and digital cameras, which allow users to associate predefined effects to a video captured by the device. As stated in our proposal, users may download effects from web-based repositories.[7]

Given the presented design, infrastructure, applications and field trial, we argue that the value of our sticker-based specialization of the watch-and-comment paradigm lies in its simplicity and generality.

Among the lessons we learned from the design, implementation and field trial reported in this paper, the following may be considered by other researchers. First, in scenarios where users are expected to be active, it is paramount that the task expected from the user adds value to the user's experience immediately. Other services may be offered *a posteriori*, as in our case with registering the moments and intervals for later review, but they will not be possible if the user does not act as expected.

Second, it is paramount that we understand how the users usually perform the tasks so that their actions can be ubiquitously included in the design.

Third, from an architectural perspective, it is essential that services and applications be integrated by means of standard infrastructures and specifications. We expect that our current implementation will facilitate the development of extensions, which can be experimented in other platforms, such as the Web, using other standards, e.g. SMIL, and in other environments.

Future work includes working on the limitations of our current proof-of-concept application. We intend to design a new application which allows the configuration

[7] An example is http://www.creative.com/mylivecam/livecentral/downloads.aspx

of sets of stickers. Moreover, to allow the current prototype to save and share the discriminated moments and intervals of interest, we plan to update some of our previous work [52] with respect to our current architecture. We plan to experiment new modules supporting the Zeroconf protocol,[8] which should leverage the use of a wide range of devices for the discrimination of media moments and intervals. We also plan to formalize an API corresponding to our current implementation to facilitate the building of new applications. Attention will also be given to studies currently demanded in terms of new qualitative evaluations that consider groups of users in novel scenarios.

Acknowledgements We thank the several agencies which provide funds to our research: FAPESP, CAPES, CNPq, FINEP, and MCT/CTIC. We also thank Rudinei Goularte, Renan G. Cattelan, Luiz F. G. Soares, Dick C. A. Bulterman and Pablo S. César for many suggestions and challenging questions. We also thank the anonymous reviewers for their suggestions and advice.

Appendix: NCL documents used in experiment

The NCL program in Listing 2, equivalent to the one presented in Listing 1, corresponds to a sticker presenting the Corinthians logo and anthem for 15 s.

```
00
00 <!-- Listing 2: soccer-Corinthians.ncl
00                 sticker with media/Corinthians.png + anthem-Corinthians.mp3
00
01 <ncl id="soccer" xmlns="http://www.ncl.org.br/NCL3.0/EDTVProfile">
02   <head>
03     <regionBase>
04       <region id="rgSticker" right="10px" top="10px"
                                width="150px" height="150px" zIndex="2"/>
05       <region id="rgAnthem"/>
06     </regionBase>
07     <descriptorBase>
08       <descriptor id="dSticker" region="rgSticker" explicitDur="15s"/>
09       <descriptor id="dAnthem" region="rgAnthem" explicitDur="15s"/>
10     </descriptorBase>
11   </head>
12     <body id="main">
13       <port id="pImg" component="img"/>
14       <port id="pAnthem" component="anthem"/>
15       <media id="img" src="media/Corinthians.png" descriptor="dSticker"/>
16       <media id="anthem" src="media/anthem-Corinthians.mp3" descriptor="dAnthem"/>
17     </body>
18 </ncl>
```

The NCL code in Listing 3 is equivalent to the one used in the original broadcast.

```
00
00 <!-- Listing 3: original interactive program
00
01 <ncl id="program" xmlns="http://www.ncl.org.br/NCL3.0/EDTVProfile">
02   <head>
03     <regionBase>
04        <region id="rgVideo" width="100
05     </regionBase>
06     <descriptorBase>
07        <descriptor id="dVideo" region="rgVideo"/>
08     </descriptorBase>
```

[8] http://www.zeroconf.org

```
09     <connectorBase>
10        <importBase documentURI="ConnectorBase.ncl" alias="conn"/>
11     </connectorBase>
12   </head>
13   <body>
14      <port id="pVideo" component="video"/>
15      <media id="video" src="media/video.mp4" descriptor="dVideo"/>
16   </body>
17 </ncl>
```

The NCL program in Listing 4, edited on the fly on the set-top-box, extends the program in Listing 3 to allow a user to add the sticker (Listing 2) while watching the match. The presentation of the sticker is activated when the user presses the red button in the remote control, and may be canceled by the user pressing the blue button.

It is interesting to observe than the new document is simple, as it should be, and that the temporal consistency is maintained, as expected [32].

```
00
00 <!-- Listing 4: (live) edited interactive program importing sticker from Listing 2
00
01 <ncl id="program" xmlns="http://www.ncl.org.br/NCL3.0/EDTVProfile">
02    <head>
03        <importedDocumentBase>
04            <importNCL documentURI="soccer-Corinthians.ncl" alias="sticker"/>
05        </importedDocumentBase>
06        <regionBase>
07            <region id="rgVideo" width="100
08        </regionBase>
09        <descriptorBase>
10            <descriptor id="dVideo" region="rgVideo"/>
11        </descriptorBase>
12        <connectorBase>
13            <importBase documentURI="ConnectorBase.ncl" alias="conn"/>
14        </connectorBase>
15    </head>
16    <body>
17        <context id="ctxSticker" refer="sticker#main"/>
18        <port id="pVideo" component="video"/>
19        <media id="video" src="media/video.mp4" descriptor="dVideo"/>
20        <link xconnector="conn#onKeySelectionStart">
21            <linkParam name="keyCode" value="RED"/>
22            <bind role="onSelection" component="video" />
23            <bind role="start" component="ctxSticker" />
24        </link>
25        <link xconnector="conn#onKeySelectionAbort">
26            <linkParam name="keyCode" value="BLUE"/>
27            <bind role="onSelection" component="video" />
28            <bind role="abort" component="ctxSticker" />
29        </link>
30    </body>
31 </ncl>
```

References

1. ABNT-2007: ABNT NBR 15606-2 Associação Brasileira de Normas Técnicas. Digital Terrestrial Television Standard 06: Data Codification and Transmission Specifications for Digital Broadcasting, Part 2 - GINGA-NCL: XML Application Language for Application Coding. http://www.abnt.org.br/imagens/Normalizacao_TV_Digital/ABNTNBR15606-2_2007Ing_2008.pdf (version 2008)
2. Allen JF, Hayes PJ (1990) Moments and points in an interval-based temporal logic. Comput Intell 5(4):225–238. doi:10.1111/j.1467-8640.1989.tb00329.x

3. Athanasiadis E, Mitropoulos S (2010) A distributed platform for personalized advertising in digital interactive TV environments. J Syst Softw 83:1453–1469. doi:10.1016/j.jss.2010.02.040
4. Blanco-Fernández Y, Arias JJP, Gil-Solla A, Cabrer MR, Nores ML, Duque JG, Vilas AF, Redondo RPD, Muñoz JB (2008) An MHP framework to provide intelligent personalized recommendations about digital TV contents. Softw Pract Exp 38(9):925–960
5. Blanco-Fernández Y, Pazos-Arias JJ, Gil-Solla A, Ramos-Cabrer M, López-Nores M (2008) Zaptv: Personalized user-generated content for handheld devices in dvb-h mobile networks. In: Proceedings of the European conference on interactive television (EUROITV'08) (LNCS 5866). Springer, Berlin, pp 193–203. doi:10.1007/978-3-540-69478-6_26
6. Bulterman DCA, Jansen AJ, César P, Mullender S, Hyche E, DeMeglio M, Quint J, Kawamura H, Weck D, Pañeda XG, Melendi D, Cruz-Lara S, Hanclik M, Zucker DF, Michel T (2008) Synchronized Multimedia Integration Language (SMIL 3.0). http://www.w3.org/TR/SMIL3/
7. Cattelan RG, Santos FS, Goularte R, Teixeira CAC, Pimentel MdGC (2008) Watch-and-comment as a paradigm toward ubiquitous interactive video editing. ACM Trans Multimed Comput Comm Appl (ACM TOMCCAP) 4(4):1–24
8. César P, Bulterman DCA, Jansen AJ (2006) The ambulant annotator: empowering viewer-side enrichment of multimedia content. In: Proceedings of the ACM symposium on document engineering (DOCENG'06). ACM, New York, pp 186–187 doi:10.1145/1166160.1166209
9. César P, Bulterman DCA, Jansen AJ (2008) Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. In: Proceedings of the European conference on interactive television (EUROITV'08) (LNCS 5866), pp 168–177
10. César P, Bulterman DCA, Jansen AJ (2009) Leveraging user impact: an architecture for secondary screens usage in interactive television. Multimedia Syst 15(3):127–142
11. César P, Bulterman DCA, Jansen AJ, Geerts D, Knoche H, Seager W (2009) Fragment, tag, enrich, and send: Enhancing social sharing of video. ACM Trans Multimed Comput Comm Appl (ACM TOMCCAP) 5(19):1–19:27. doi:10.1145/1556134.1556136
12. Chi MC, Yeh CH, Chen MJ (2009) Robust region-of-interest determination based on user attention model through visual rhythm analysis. IEEE Trans Circuits Syst Video Technol 19:1025–1038. doi:10.1109/TCSVT.2009.2022822
13. Coppens T, Trappeniers L, Godon M (2004) AmigoTV: Towards a social TV experience. In: Proceedings of the European conference on interactive television (EUROITV'04)
14. Costa M, Correia N, Guimarães N (2002) Annotations as multiple perspectives of video content. In: Proceedings of the ACM international conference on multimedia (MULTIMEDIA'02), pp 283–286. doi:10.1145/641007.641065
15. Costa RMdR, Moreno MF, Rodrigues RF, Soares LFG (2006) Live editing of hypermedia documents. In: Proceedings of the ACM symposium on document engineering (DOCENG'06). ACM, New York, pp 165–172. doi:10.1145/1166160.1166202
16. de Ávila PM, Zorzo SD (2009) Recommender TV - A Personalized TV Guide System Compliant with Ginga. In: Proceedings of the international conference on security and cryptography (SIGMAP'09), part of the international joint conference on e-business and telecommunications (ICETE'09), pp 149–156
17. de Freitas GB, Teixeira CAC (2009) Ubiquitous services in home networks offered through digital TV. In: Proceedings of the ACM symposium on applied computing (SAC'09). ACM, New York, pp 1834–1838. doi:10.1145/1529282.1529691
18. De Lucena VF, Filho JEC, Viana NS, Maia O (2009) A home automation proposal built on the ginga digital TV middleware and the OSGi framework. IEEE Trans Consum Electron 55(3):1254–1262
19. Deigmoeller J, Itagaki T, Stoll G, Just N (2010) An approach to intelligently crop and scale video for broadcast applications. In: Proceedings of the ACM symposium on applied computing (SAC'10). ACM, New York, pp 1911–1918. doi:10.1145/1774088.1774493
20. Di Massa R, Montagnuolo M, Messina A (2010) Implicit news recommendation based on user interest models and multimodal content analysis. In: Proceedings of the international workshop on automated information extraction in media production (AIEMPro'10). ACM, New York, pp 33–38. doi:10.1145/1877850.1877861
21. Dimitrova N, Janevski A, Li D, Zimmerman J (2003) Who's that actor?: the Infosip TV agent. In: Proceedings of the ACM SIGMM workshop on experiential telepresence (ETP'03). ACM, New York, pp 76–79. doi:10.1145/982484.982499
22. Drucker SM, Glatzer A, De Mar S, Wong C (2002) Smartskip: consumer level browsing and skipping of digital video content. In: Proceedings of the ACM conference on human factors in computing systems (CHI'02). ACM, New York, pp 219–226. doi:10.1145/503376.503416

23. Furini M, Geraci F, Montangero M, Pellegrini M (2010) Stimo: Still and moving video storyboard for the web scenario. Multimedia Tools and Applications 46:47–69. doi:10.1007/s11042-009-0307-7
24. Gao Y, Dai QH (2008) Clip based video summarization and ranking. In: Proceedings of the international conference on content-based image and video retrieval (CIVR'08). ACM, New York, pp 135–140. doi:10.1145/1386352.1386375
25. Goularte R, Cattelan RG, Camacho-Guerrero JA, Inácio VR Jr, Pimentel MdGC (2004) Interactive multimedia annotations: enriching and extending content. In: Proceedings of the ACM symposium on document engineering (DOCENG'04), pp 84–86. doi:10.1145/1030397.1030414
26. Hickson I (2011) HTML5 a vocabulary and associated APIs for HTML and XHTML—W3C Working Draft 13 January 2011. http://www.w3.org/TR/html5/
27. Hölbling G, Rabl T, Coquil D, Kosch H (2008) Interactive TV services on mobile devices. IEEE Multimed 15(2):72–76. doi:10.1109/MMUL.2008.34
28. Hsu SH, Wen MH, Lin HC, Lee CC, Lee CH (2007) AIMED–a personalized TV recommendation system. In: Proceedings of the European conference on interactive television (EUROITV'07) (LNCS 4417), pp 166–174
29. Huang EM, Harboe G, Tullio J, Novak A, Massey N, Metcalf CJ, Romano G (2009) Of social television comes home: a field study of communication choices and practices in TV-based text and voice chat. In: Proceedings of the ACM international conference on human factors in computing systems (CHI'09), pp 585–594
30. Huet B, Jiten J, Merialdo B (2005) Personalization of hyperlinked video in interactive television. In: Proceedings of the IEEE conference on multimedia and expo, p 4. doi:10.1109/ICME.2005.1521460
31. ISO (1996) Information technology—generic coding of moving pictures and associated audio information: digital storage media command & control. ISO/IEC 13818-6
32. Jansen J, César P, Bulterman DC (2010) A model for editing operations on active temporal multimedia documents. In: Proceedings of the ACM symposium on document engineering (DOCENG'10). ACM, New York, pp 87–96. doi:10.1145/1860559.1860579
33. Laiola Guimarães R, César P, Bulterman DC (2010) Creating and sharing personalized time-based annotations of videos on the web. In: Proceedings of the ACM symposium on document engineering (DOCENG'10). ACM, New York, pp 27–36. doi:10.1145/1860559.1860567
34. Liu Z, Gibbon DC, Drucker H, Basso A (2008) Content personalization and adaptation for three-screen services. In: Proceedings of the international conference on content-based image and video retrieval (CIVR '08). ACM, New York, pp 635–644. doi:10.1145/1386352.1386444
35. Liu Z, Zavesky E, Shahraray B, Gibbon D, Basso A (2008) Brief and high-interest video summary generation: evaluating the at&t labs rushes summarizations. In: Proceedings of the ACM TRECVid video summarization workshop (TVS'08). ACM, New York, pp 21–25. doi:10.1145/1463563.1463565
36. Lynn SG, Olsen DR Jr, Partridge BG (2009) Time warp football. In: Proceedings of the European conference on interactive television (EUROITV'09). ACM, New York, pp 77–86. doi:10.1145/1542084.1542098
37. Miyahara M, Aoki M, Takiguchi T, Ariki Y (2008) Tagging video contents with positive/negative interest based on user's facial expression. In: Satoh S, Nack F, Etoh M (eds) Advances in multimedia modeling (LNCS 4903). Springer, Berlin, pp 210–219
38. Motti VG, Fagá R Jr, Catellan RG, Pimentel MdGC, Teixeira CA (2009) Collaborative synchronous video annotation via the watch-and-comment paradigm. In: Proceedings of the European Conference on Interactive Television (EUROITV'09). ACM, New York, pp 67–76. doi:10.1145/1542084.1542097
39. Nathan M, Harrison C, Yarosh S, Terveen LG, Stead L, Amento B (2008) CollaboraTV: making television viewing social again. In: Proceedings of the international conference on designing interactive user experiences for TV and video (UXTV'08). ACM, New York, pp 85–94
40. Nitta N, Takahashi Y, Babaguchi N (2009) Automatic personalized video abstraction for sports videos using metadata. Multimedia Tools and Applications 41(1):1–25
41. Patel M, Gossweiler R, Sahami M, Blackburn J, Brown D, Knight A(2008) Google TV search: dual-wielding search and discovery in a large-scale product. In: Proceedings of the international conference on designing interactive user experiences for TV and video (UXTV'08). ACM, New York, pp 95–104. doi:10.1145/1453805.1453826
42. Petersen MK, Butkus A (2008) Modeling emotional context from latent semantics. In: Proceedings of the international conference on designing interactive user experiences for TV and video (UXTV'08). ACM, New York, pp 63–66. doi:10.1145/1453805.1453819

43. Pimentel MdGC, Cattelan RG, Melo EL, Teixeira CA (2008) End-user editing of interactive multimedia documents. In: Proceedings of the ACM symposium on document engineering (DOCENG'08). ACM, New York, pp 298–301. doi:10.1145/1410140.1410204
44. Pimentel MdGC, Cattelan RG, Freitas G, Melo EL, Teixeira CAC (2009) Watch-and-comment as an approach to collaborative annotate points of interest in video and interactive-TV programs. In: Marcus A, Roibás AC, Sala R (eds) Mobile TV: customizing content and experience. Springer, Berlin, pp 349–366
45. Pimentel MdGC, Cattelan RG, Melo EL, Prado AF, Teixeira CAC (2010) End-user live editing of iTV programmes. Int J Adv Media Comm 4(1):78–103. doi:10.1504/IJAMC.2010.030007
46. Ramos G, Balakrishnan R (2003) Fluid interaction techniques for the control and annotation of digital video. In: Proceedings of the ACM symposium on user interface software and technology (UIST'03), pp 105–114. doi:10.1145/964696.964708
47. Rogge B, Bekaert J, de Walle R (2004) Timing issues in multimedia formats: review of the principles and comparison of existing formats. IEEE Trans Multimedia 6(6):910–924
48. Schulzrinne H, Rao A, Lanphier R (1998) Real Time Streaming Protocol (RTSP). http://www.ietf.org/rfc/rfc2326.txt
49. Soares LFG, de Souza Filho (2007) Interactive television in Brazil: system software and the digital divide. In: Proceedings of the European conference on interactive television (EUROITV'07) (LNCS 4417). Springer, Berlin, pp 41–44
50. Soares LFG, Rodrigues RF, Moreno MF (2007) Ginga-NCL: the declarative environment of the Brazilian digital TV system. J Braz Comput Soc 12(4):37–46
51. Soares LF, Rodrigues RF, Cerqueira R, Barbosa SD (2010) Variable and state handling in NCL. Multimedia Tools and Applications 50:465–489. doi:10.1007/s11042-010-0478-2
52. Teixeira CAC, Freitas G, Pimentel MdGC (2010) Distributed discrimination of media moments and media intervals: a watch-and-comment approach. In: Proceedings of the ACM symposium on applied computing (SAC'10), pp 1929–1935
53. Tjondronegoro D, Chen YPP, Pham B (2004) Integrating highlights for more complete sports video summarization. IEEE MultiMed 11:22–37. doi:10.1109/MMUL.2004.28
54. Troncy R, Mannens E, Pfeiffer S, Deursen DV (2010) Media fragments URI 1.0—W3C working draft, 8 December 2010. http://www.w3.org/2008/WebVideo/Fragments/WD-media-fragments-spec/
55. Vandermolen H, Wiegering A, Sommers M, Levine SJ, Franco A (2009) Identifying events of interest within video content. http://www.freepatentsonline.com/7624416.html US Patent(7624416)
56. Vuorimaa P, Bulterman D, César, P (2008) SMIL timesheets 1.0—W3C working draft, 10 January 2008. http://www.w3.org/TR/2008/WD-timesheets-20080110/
57. Wei Y, Bhandarkar SM, Li K (2009) Client-centered multimedia content adaptation. ACM Trans Multimed Comput Comm Appl (ACM TOMCCAP) 5(22):1–22:26. doi:10.1145/1556134.1556139
58. Xie L, Sundaram H, Campbell M (2008) Event mining in multimedia streams. Proc IEEE 96: 623–647
59. Yamamoto M, Nitta N, Babaguchi N (2006) Estimating intervals of interest during TV viewing for automatic personal preference acquisition. In: Zhuang Y, Yang S, Rui Y, He Q (eds) Advances in multimedia information processing (PCM'06) (LNCS 4261). Springer, Berlin, pp 615–623

Cesar A. C. Teixeira is an Associate Professor in the Computer Science Department at the Universidade Federal de São Carlos, Brazil, where he is the head of the Laboratory for Innovation in Computer Science and Engineering. Dr. Teixeira received his PhD in electrical engineering the Universidade de São Paulo in 1989. His research interests focus on distributed multimedia systems, in particular in the context of Interactive TV. He is a member of the ACM, the IEEE, and the Brazilian Computer Society (SBC).

Erick L. Melo is a PhD canditate at the Universidade Federal de São Carlos. He received BSc and MSc degrees in Computer Science from the Universidade Federal de São Carlos. His research interests include interative multimedia and digital TV. He is member of the Brazilian Computer Society (SBC).

Giliard B. Freitas is an Assistant Professor in the Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, Campus Confresa. He received BSc and MSc degrees in Computer Science from the Universidade Federal do Tocantins and the Universidade Federal de São Carlos, respectively. His research interests include interative multimedia, mobile computing and digital TV. He is a member of the Brazilian Computer Society (SBC).

Celso A. S. Santos is an Associate Professor in the Computer Science Department at the Universidade Federal da Bahia, Brazil, where he is the head of the Deparment of Computer Science. Dr. Santos received his PhD from the Université Paul Sabatier de Toulouse III in 1999. His research interests focus on multimedia systems and Interactive TV. He is a member of the ACM, the IEEE, and the Brazilian Computer Society (SBC).

Maria da Graça C. Pimentel is a Full Professor in the Computer Science Department at the Universidade de São Paulo, Campus São Carlos, Brazil, where she is the head of the Intermdia Laboratory. Dr. Pimentel received her PhD in Computer Sciente in 1994 from the University of Kent, UK. Her research interests includes multimedia systems, ubiquitous computing and Interactive TV. She is a member of the ACM, the IEEE, and the Brazilian Computer Society (SBC).