Exploring the Use of Augmented Reality Concepts to Enhance the TV Viewer Experience

Simão Carvalho[1], Teresa Romão[1(✉)], and Pedro Centieiro[1,2]

[1] NOVA LINCS, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
Spn.carvalho@campus.fct.unl.pt, tir@fct.unl.pt, pcentieiro@gmail.com
[2] Viva Superstars Digital Media, Lda, Madan Parque, Rua Dos Inventores, 2825-182 Caparica, Portugal

Abstract. Television has no longer the same effect on viewers as it had decades ago. The time that was formerly spent watching "traditional" television is now shared with or replaced by mobile devices, such as smartphones and tablets. When using these devices, the viewer has the opportunity to further interact with the content that is provided to him, as well as with remote friends. The work presented in this paper explores new concepts of interaction in television contexts, which comprises the integration of augmented reality (AR) techniques with television shows to enhance the viewer's experience, allowing them to access additional information about the content they are watching and to create their own content to share with their friends. This paper describes the developed prototype and the corresponding preliminary user evaluation with promising results.

Keywords: Augmented reality · TV shows · Interaction with TV content · Social interaction · Mobile devices · Entertainment

1 Introduction

Once available to the world's population, television became an essential part of our daily routine. This technology can show us news from around the world and it can be a rich source of entertainment. This made television a main form of socializing, as well as a way to access information [1] until recently, when some changes in this scenario started to happen, as mobile devices started to gain ground against the TV, especially among young adults [2]. Younger generations are more familiar with interactive technology which provides them with the content they want at anyplace and at anytime. As smartphones and tablets can do the same as traditional television and much more, television is losing the audience it used to have among the younger generations.

Since its creation to the present day, television has undergone different evolutions to adapt to the new generations of viewers (such as the introduction of TV set boxes and IPTV). It is now possible to watch television over the Internet on a mobile device, and since it is estimated that more than 63% of world's population is in possession of a

C. Sylla and I. Iurgel (Eds.): TIE 2019, LNICST 307, pp. 55–64, 2020.
https://doi.org/10.1007/978-3-030-40180-1_6

mobile phone [3], television is on the verge of another major evolution. Like we use social networks to interact with multimedia content and to communicate and socialize with remote friends, we aim to bring the same kind of interactivity to the traditional passive experience of watching television.

This way, we seek to contribute to turn watching TV into a more interactive activity of entertainment and socialization. Thus, we have been exploring new forms of interaction with TV content that enhance the viewers' experience. To facilitate the viewers' access to additional information related with the TV show content they are watching, we explored the integration of Augmented Reality (AR) concepts with TV shows' content and developed tools for TV content edition and share. In this paper, we present our first concept, its realization as a prototype and results of usability tests which attempt to validate our approach and uproot design flaws.

2 Related Work

AR allows the user to observe the real world with superimposed virtual objects in a way that they seem to co-exist in the same space as real world objects. When we use AR, we intend to improve user's perception of and interaction with the real world [4], providing him with additional information about his surrounding world. From head-mounted display based to mobile applications, AR applications have been developed for a variety of areas, such as health [5], environmental management [6], education [7], tourism [8] and entertainment [9]. AR has also been used for real time augmentation of broadcast video, to enhance the visualization of sporting events and to place advertisements in the scene. An early example is the FoxTrax system, which tracks and highlights the location of a hockey puck as it moved rapidly across the ice [10]. In these situations, the end-viewer is not able to control the information he wants to visualize. There are already some studies in the area of mixing augmented reality with television broadcast, where the viewer can interact with the show he is watching. In [11], the authors propose a system that enables viewer-selectable augmented broadcasting services with the need for an extra device.

Interactive television combines traditional TV services with data services, allowing users' participation and feedback. It allows various forms of interaction, such as with TV-related content and TV services, delivered through specific pay-TV set-top boxes and controlled by extra devices, such as remote controls, mobile phones or tablets. Users may choose the viewing angle thanks to the existence of various cameras filming a match, participate in trivia quizzes along with other participants in the studio or at their own homes or check statistics. Olsen et al. [12] presents a system for interactive television news and the corresponding evaluation to understand interactive viewing behavior.

To provide the viewers with a more active role, we have been exploring how to create more interactive experiences, that allow them to control what they want to watch, when and where, as well as to access additional information and to create and share content based on the TV show they are watching. Thus, as explained next, we

integrated AR techniques with video playback to complement the TV show content, and propose mechanisms for user social interaction that allows them to edit the video content they are watching and share the outcomes with their friends. Unlike, second screen applications for interaction with TV content [13], using our prototype the interaction is achieved through a unique mobile device (tablet or smartphone) that most users own and can carry wherever they want.

3 TvTeller: Prototype Description

We developed TvTeller prototype as a proof of concept which aims at exploring and changing the way people interact with the TV content they are watching. Using a mobile device as the screen for visualization and interaction, viewers are able to access additional information about the TV content they are watching, as well as to extract, edit and share video clips. This way, TvTeller provides users with the possibility to know more about TV series characters, create amusing content, and share it with their friends. As users, especially young people, often interact with their mobile devices to share content with friends and search for information, TvTeller explores how to non-intrusively integrate these activities within the experience of TV watching.

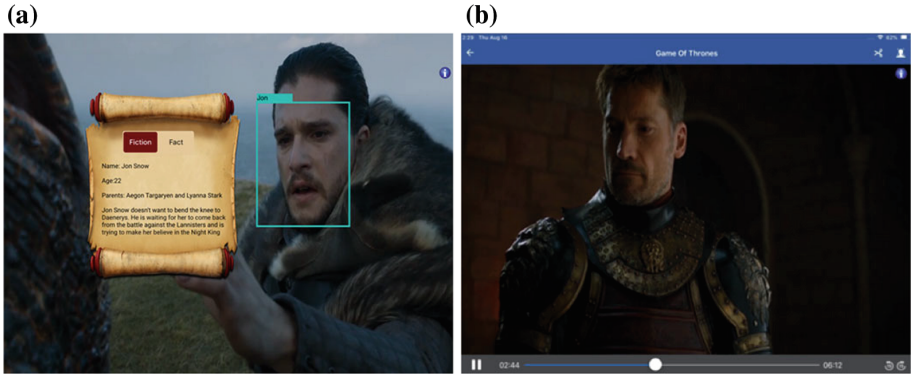

Fig. 1. While watching a TV series episode the user can click on a character to know more about him (a) when a character is identified for the first time and (b) when a character is recognized again later on (an icon "i" appears).

TvTeller is composed by two components that are independent from each other. One is more information-oriented and the other more entertainment-oriented. The first one allows the user to select a TV show episode and, while watching it, to know more about the characters or some other important elements, which may include animals and objects that appear on screen while the episode is played. While watching traditional television, when a user wants to know more about what he is watching, he needs to

search for it outside the TV context (for example, using his mobile phone to search the Internet), diverting his attention from the TV show. Using TvTeller, the TV series is augmented with virtual information objects superimposed on the video which provide additional information regarding the elements of the displayed scene (Fig. 1a). The user does not need to move his eyes from the screen to obtain information on a new TV series he just started to follow or recall details from a TV series former season. Instead of superimposing virtual objects on the video coming from a camera, as in "traditional" mobile AR systems, we superimpose the virtual objects directly on the video itself.

The entertainment-oriented component allows users to trim parts of an episode, creating short video clips that can be edited and shared later on. This is a non-intrusive task, as the viewer can trim a moment that he liked while he continues to watch the episode (as explained in Sect. 3.1). After the trim is done, the user can edit the trimmed video adding virtual objects to the characters that appear in it (Fig. 2) and share the creations with others. Since sharing funny content is something very usual nowadays (e.g. meme photos and videos), we intend to facilitate this task and enhance the user's motivation to watch the episode with more attention, as he can find and record the perfect moment(s) of the episode to share.

Fig. 2. Gallery editor view: (a) adding virtual objects to a character and (b) adding a bubble object.

Although this concept may be applied to any TV series or show, the series "Game of Thrones" was used as the case study. This series, due to its popularity and its unique characteristics, is a good example to demonstrate our concept. For example, it has a lot of characters, so it is easy to forget who someone is and whether he did something that changed the course of the plot. Besides, it is a show with some recasts, which can confuse a viewer when he sees a new actor for the first time playing an existing role. Our concept can also be used in movies (especially when they have sequels), as well as with almost every TV content, such as reality shows, sitcoms, or documentaries.

3.1 Interaction

In the TvTeller main screen, the user can choose from watching a TV series episode or go to the gallery where he can find the clips that he has saved earlier.

When the user selects to watch an episode, he is presented with a grid view of the available series from which he can select the one he wants to watch. After selecting an episode, the video starts playing and the user can activate the recognition feature, touching on the face button on the top right of the screen in the navigation bar (Fig. 1b). When this feature is activated, the faces of the characters on the screen are recognized and a rectangular bounding box appears around them when a character is identified for the first time (Fig. 1a). The following times, only an icon "i" appears on the top right of the screen, in a less intrusive way (Fig. 1b). The user can touch on the character to visualize information about the character/actor. It is also possible to access information about non-human characters or objects. Since these elements have no face, to inform the user that he can access more information about the character or the object, an icon "i" also appears on the top right of the screen.

Besides accessing information regarding the elements on the screen while watching a series, when the user likes a specific moment of the episode, he can trim that moment of the video which is saved in the gallery for later edition and sharing. In order to facilitate this task (video trimming), the user only has to touch on the trim button, which looks like a scissor, when he watches the moment he wants to save. 15 s of the video are recorded: 13 s before the button touch (as the user decides to trim the video after he visualized the moment) and 2 s after (preventing the desired clip to end abruptly). We chose this duration because this kind of clips should be kept short and, according to our experiments, in most situations, this duration is appropriate to capture the relevant content the user wants to record.

In the gallery, the user can find all the moments he saved while watching his series. Selecting one of these moments (video clip), opens it in the editor view (Fig. 2). Here, the user is able to choose one of the available virtual objects to superimpose on the characters and change their appearance. There are two types of objects: face objects and bubble objects. Face objects, such as moustaches or glasses, are automatically superimposed on characters' faces, adjusting to the face layout. Bubble objects, such as thought balloons, are placed, by the user, near the face of a specific character. After the user drags the thought balloon image close to a character face, he touches on it to bound the object to the face. The pre-defined bubble objects appear near the selected face for 1.5 s (Fig. 2b). In the empty thought balloon, the user needs to write the text to be displayed. This balloon appears for 3 s instead of 1.5 s, to give the users time to read the information in it. After adding an object, the user can remove the objects if he does not like them, add more objects, save the video in the gallery or share his creation.

3.2 Implementation and Architecture Workflow

Three main algorithms were developed to enable the TvTeller operation (Fig. 3).

To avoid performance issues from deteriorating the user experience while watching a TV show, each TV show episode must be pre-processed before it is ready to be watched. The facial and object recognition algorithm (1) receives the TV show file as

the input and returns a JSON file containing the characters' faces and objects that appear in the frames of the video. Every 0.4 s (achieved by experimental observation), a frame of the video is analyzed and faces and objects are recognized. When recognition is achieved with more than 80% confidence, the identity of the face (or object), its location on the frame and the time of the frame are stored. This information, stored in the output JSON file, is used to provide the user with the augmented information about the characters and objects while he is watching the TV show, giving him the impression that it is made in real time. Detection and recognition are done using algorithms from Apple's Vision framework [14]. Apple has already some machine learning models to identify a variety of objects, but since we need some very specific models, we had to create them ourselves, using Tensorflow [15]. Another JSON file that contains the information about the characters and works as a dictionary, linking the ids of the recognized characters with the corresponding information, is part of our bundle resources. When the user taps the screen to watch additional information about a recognized character or special element, a function fetches this information from the dictionary.

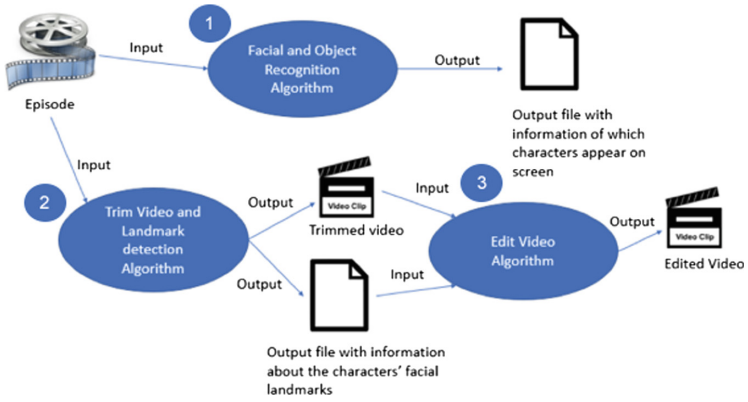

Fig. 3. Algorithms used in TvTeller.

To allow users to augment the video clips they have trimmed and stored in the gallery with virtual objects that must be registered over the characters' faces (face objects), an algorithm for face landmark detection (2) was developed. This algorithm detects the facial landmarks of every character that appears on the trimmed video and runs right when the user clicks to trim the video while watching an episode. First, the video is trimmed and saved inside the app's document directory. After that, a function runs the landmark detection algorithm and produces a JSON file containing the landmarks positions. We need to know with a good precision where are the landmarks located, so we can precisely superimpose objects on top of a face or near it. The augmented objects need to move with the character as smoothly as possible along the

video clips, so the ideal was to gather information related to all frames of the trimmed video. However, this was time-consuming, so after some experimentation, we confirmed that analyzing a frame every 0.1 s provides sufficient information to achieve an appropriate dynamic registration of the augmented objects along the whole video in a suitable time period.

The edit video algorithm (3) is responsible for overlaying the selected virtual objects to all the frames in a trimmed clip stored in the gallery that is being edited and for the creation of the new video that can be shared. This algorithm uses the output file generated by the landmark detection algorithm (2) described before. After the user places an object on the clip being edited, the algorithm starts by identifying the nearest face in the frame. Then, it checks the position of the face in the next frame 0.1 s latter (this is done to move the virtual object smoothly between frames, in order to follow the face). If the face is near the same position, then the algorithm starts using the new position to check for the position of the face in the following frame (also in the next 0.1 s), otherwise (or if the face is not detected during 0.5 s) the virtual object is removed. This process goes on until the face disappears or for the maximum time an object is kept on the clip.

4 Evaluation

After developing the TvTeller prototype, user tests were performed to evaluate its usability, test our concept and to gather feedback regarding how the application was going to be used, so we could fix some flaws and make it more user friendly.

4.1 Participants and Methodology

Users tests were conducted with a total of 30 users (22 were male and 8 were female) with ages between 14 and 53 years old ($\bar{x}$ = 26,7). All tests were made using the same device, an iPad Air 2. Before starting the tests, the users were provided with a brief explanation of the TvTeller concept, and during the tests, we observed the behaviour of the participants, took notes, and provided assistance if anyone had a problem. The participants were provided with a list of tasks to execute, available on a computer, but they were free to use the prototype as they liked. The tasks included watching an episode from Game of Thrones, turning character recognition on and visualizing augmented information about characters, actors and objects, trimming a moment of the episode, editing the trimmed video clip adding a bubble object, and sharing the resulting clip on a social network. After the test, the users filled in a questionnaire.

4.2 Results and Discussion

The questionnaire focused on usability and user experience issues, including general feedback about the TvTeller, its usability and ease of use. The first section included participants' personal data, such as age and genre. The second section contained

questions about their habits while watching TV series. Half of the participants (50%) watch series on a daily basis, 13.3% watch almost every day, 30% watch series on a weekly basis (some users stated that this is due to the series they watch only released one episode per week) and 6.7% rarely watch TV series. The most used devices to watch TV series are: the computer selected by 83.3% of the participants, television (70%), smartphone (36.7%) and tablets (6.7%). Finally, 90% of the participants stated they use to watch meme videos, meaning that they were familiar with the theme.

The third section concerns how easy it was to perform the main available tasks. The users were asked to rate statements, using a five-point Likert-type scale, which ranged from "Very hard" (1) to "Very easy" (5). Table 1 shows that most of the tasks were considered very easy to perform by most of the users (S1–S3, S5). Adding objects to the trimmed video on the gallery was not so easy, as users were trying to understand how the face and bubble objects worked (S4). Some suggested to change the face objects' images, because the ones used give the impression that not only a moustache or glasses will be added over the characters' face, but the whole smiley image.

Table 1. Summary of the questionnaire results (third section). Higher scores are highlighted.

Statements	Very Hard	Hard	Neutral	Easy	Very Easy
S1. Activate facial recognition	0%	0%	10%	16.7%	73.33%
S2. Extract a clip from the video	0%	0%	3.33%	30%	66.67%
S3. Visualize info about the characters	0%	0%	0%	30%	70%
S4. Add objects to trimmed video clips	0%	3.3%	6.67%	46.67%	43.33%
S5. Share the clip on a social network	0%	0%	0%	13.33%	86,67%

The fourth section focuses on usability and entertainment aspects. In this section, the users were asked to rate statements, using a five-point Likert-type scale, which ranged from "Strongly Disagree" (1) to "Strongly Agree" (5), as presented in Table 2. All users considered the augmented information about the characters useful and complementary for their experience (S6). The characters' facial recognition bounding box (Fig. 1a) did not disturb the experience of most users, but a few users complained (S7). Most users were able to trim video clips without losing their focus on the series they were watching (S8), and they stated they would use the facial recognition feature in a real-life context (S9). While a few users were not interested in creating new video content (S10), most users said they would react to content shared by others (S11) and considered they could have fun watching that content (S12). Users also stated that TvTeller may help them to catch up on a series they have not follow for a while or may help them to follow an episode when they are performing another task while watching it. Observation of the users while testing the prototype corroborates with these results.

Table 2. Summary of the questionnaire results (fourth section). Higher scores are highlighted.

Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
S6. Additional information about the characters is useful and complemented my experience	0%	0%	0%	30%	70%
S7. The appearance of the facial recognition box did not disturb my experience	6.67%	3.33%	10%	43.33%	36.67%
S8. I can easily extract a video clip without losing focus on the series I am watching	0%	6.67%	13.33%	30%	50%
S9. In a real context, I would use facial recognition functionality while watching my series	0%	6.67%	23.33%	30 %	40%
S10. Probably, I will create content to share	3.33%	13.33%	23.33%	33.33%	26.67%
S11. Probably, I will react to content created by others	0%	0%	16.67%	46.67%	36.67%
S12. Even if I do not watch a show, I can entertain myself with content that others created	0%	3.33%	13.33%	33.33%	50%

5 Conclusions and Future Work

With this work we aimed at exploring the possibility of turning the experience of watching a TV show into a more active and social one, allowing users to non-intrusively perform tasks they usually carry out while watching TV, such as search for information or share content with friends. In order to evaluate our concept, we created a prototype that gives the users the possibility to know more about their favorite series resorting to facial recognition and the opportunity to create new content to share with friends using augmented reality concepts.

In general, users liked the features presented in this prototype and the possibility of effortlessly perform additional tasks that do not disturb their focus on the TV content. Adding objects to a face, such as the moustache, is something that they are already used to, using face filters on applications like Instagram or Snapchat, and they loved the possibility to add them to TV characters. The feature that allows users to visualize information about the characters was also very well received. Some users stated that this feature solves some of their problems while watching a TV series, since they had the necessity to know more about the characters while they were watching TV series. Most of the users were able to perform the tasks without losing their focus on the TV content, however improvements can still (and will) me made.

In the future, we aim to solve some problems reported by users, regarding the UI and the facial detection algorithm. We also aim to take this concept even further and identify the facial expression of the characters to create new animations based on emotions, add sound clips or change the characters' faces instead of only adding objects to them.

Acknowledgments. This work is funded by FCT/MCTES NOVA LINCS PEst UID/CEC/04516/2019.

References

1. Gerbner, G., Gross, L., Morgan, M., Signorielli, N.: Living with television: the dynamics of the cultivation process. In: Bryant, J., Zillman, D. (eds.) Perspectives on Media Effects, pp. 17–40. Lawrence Erlbaum, New Jersey (1986)
2. Richter, F.: Smartphones beat Tv for young adults in the U.S, Statista (2017). https://www.statista.com/chart/8660/smartphone-vs-tv-usage/. Accessed 20 June 2019
3. Statista.: Number of mobile phone users worldwide from 2013 to 2019 (in billions). https://www.statista.com/statistics/274774/fore-cast-of-mobile-phone-users-worldwide/. Accessed 20 June 2019
4. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. Appl. **21**(6), 34–47 (2001)
5. Bernhardt, S., Nicolau, S.A., Soler, L., Doignon, C.: The status of augmented reality in laparoscopic surgery as of 2016. Med. Image Anal. **37**, 66–90 (2017)
6. Romão, T., et al.: ANTS – augmented environments. Comput. Graph. **28**(5), 625–633 (2004)
7. Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook: a transitional AR interface. Comput. Graph. **25**(5), 745–753 (2001)
8. Střelák, D., Škola, F., Liarokapis, F.: Examining user experiences in a mobile augmented reality tourist guide. In: Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA 2016), Corfu, Greece, 29 June–01 July. ACM New, York (2016)
9. Lv, Z., Halawani, A., Feng, S., Ur Réhman, S., Li, H.: Touch-less interactive augmented reality game on vision-based wearable device. Pers. Ubiquit. Comput. **19**(3–4), 551–567 (2015)
10. Cavallaro, R.: The FoxTrax hockey puck tracking system. IEEE Comput. Graph. Appl. **17** (2), 6–12 (1997)
11. Kim, S.C., Koo, H.S., Kim, H., Cheong, J.: Implementation of AR-based hybrid broadcasting system by TV viewer's preferred content provider. In: Proceedings of Information Science and Security (ICISS 2016), Pattaya, Thailand, 19–22 December. IEEE (2016)
12. Olsen, D.R., Sellers, B., Boulter, T.: Enhancing interactive television news. In: Proceedings of TVX 2014, Newcastle, UK, 25–17 June. ACM, New York (2014)
13. Centieiro, P., Cardoso, B., Romão, T., Dias, A.E.: If you can feel it, you can share it! A system for sharing emotions during live sports broadcasts. In: Proceedings of ACE 2014, Funchal, Portugal, 11–14 November. ACM, New York (2014)
14. Apple Vision framework: https://developer.apple.com/documentation/vision. Accessed 06 Feb 2019
15. Tensorflow: https://www.tensorflow.org/. Accessed 06 Apr 2019