Use Second Screen to Enhance TV Viewing Experiences

Yu-Ling Chuang, Chia-Wei Liao, Wen-Shiuan Chen, Wen-Tsung Chang, Shao-Hua Cheng, Yi-Chong Zeng, and Kai-Hsuan Chan

Institute for Information Industry
1F., No. 133, Sec. 4, Minsheng E. Rd., Taipei City, Taiwan, R.O.C.
{ilmachuang,cliao,clairechen,wtchang,briancheng,
yichongzeng,kaihsuanchan}@iii.org.tw

Abstract. In this paper, we propose a second-screen interactive TV viewing scenario, and build an Android-based prototype to verify our concept and design. The demo app detects the TV program that a user is watching, and then dynamically pushes additional information through the second screen to the user. In this prototype, we apply audio fingerprinting technology to detect TV programs, and users do not have to manually surf hundreds of channels. In addition, face recognition and tennis event detection technologies have been employed to extract critical elements from the TV contents. With these content analysis tools, we can obtain video metadata more easily and need not rely on content providers. A usability test with 20 participants has been conducted. Our test shows participants are interested in our second screen app, and they would like more features and information from the app. Next, we plan to further refine the app's design to enhance user experience.

Keywords: Second Screen, User Experience, Video Content Analysis, Audio Fingerprinting.

1 Introduction

The survey of Red Bee Media shows that 52% of the respondents would like to search for additional information about the programs they have seen [1]. Unfortunately, most of the information on the Internet is fragmented. As the mobile device gets more popular, people use it as a second screen while watching TV. In Nielsen's 2011 report, 42% of tablet owners and 40% of smartphone owners use their devices in the TV viewing context [2]. For now, there is no direct link between TV and the second screen which are otherwise tightly coupled. Mobile devices such as smart phones and pads, as second screens, meet the increasing viewers' demand for information. We synchronize TV viewers' mobile devices with the TV, and push related real-time information to their devices (such as details of the products placed in the show, introduction of the cast). Although interactive TV comes with similar capability, interactive applications on the TV screen has been proved interruptive [3]. Using mobile devices as second screens of TV, our approach enhances the TV viewing experiences without much distraction to viewers. Moreover, second screen approach

P.L.P. Rau (Ed.): CCD/HCII 2013, Part I, LNCS 8023, pp. 366–374, 2013.

offers a far richer interaction mechanism than conventional remote control devices coming with interactive TV [4].

Since product placement is restricted in certain areas (such as Taiwan), second screen concept provides another channel for TV marketing. We have built a prototype based on this concept to demonstrate and validate this second-screen approach with several pre-defined scenarios.

Our prototype is an Android application. It detects the TV program being watched, and pushes real-time information accordingly. To make sure that our design satisfies users' needs while watching different shows, we include a variety of genres in the prototype. Considering the needs of viewers and the capability of technology, we push primarily the brief information like introduction, summaries, and additional contents about the program. We collect the above information and timestamp manually unless they can be obtained from the content owner. However, to be more independent of content providers in the future, automatic video tagging and annotation based on, say, face recognition and sport event detection technologies have been developed to analyze TV programs and events in tennis games. Tags and annotations are, in our case, the metadata, and will later be pushed to viewer's second screen. To automatically synchronize the mobile devices with TV, we have also developed audio fingerprint detection and matching techniques to automatically identify the TV program.

There are 20 participants in our experiment to test the prototype, and each participant uses a second screen while watching TV in our simulation. During the experiment, they need to finish some assigned tasks and then give their feedback. This is to know participants' preferences when using a second screen and the information they would like most.

2 Related Work

Fallahkhair et al [5~6] combined interactive TV and mobile phones to facilitate informal language learning. Whoever wants to study a foreign language can learn from TV shows in that language. The researchers tried to extend this approach from TV to mobile devices. Viewers activate TAMALLE, running on interactive TV and mobile phones, to get instant information of exotic culture or help with language items before, during, or after watching the show. Language items can be saved for future use. Mobile phones also provide an alternative way of learning on the move.

In 2011, Basapur et al [7] proposed a dual device TV viewing experience. They built a browser-based web application, which updated time-synchronized feeds of the program. In their prototype, a browser tab shows content related post, comments from social network, and related multimedia. Another one shows episode and show details, and actor bios. Users need to click on a specific post to open its original source website in a new tab. In this prototype, researchers manually created the feeds within 24 hours after the program airs on the TV. A year later, they built another prototype [8] where the feeds came from the viewers' social circles instead of experts, so they could interact with the prototype during a live TV show. Thus these information feeds should be more personal and more relevant to program contents. However, no

automatic synchronization mechanism has been deployed in these two prototypes. Viewers have to manually perform synchronization between devices.

Nowadays, more and more second screen apps are available on the market, such as IntoNow [9], GetGlue [10], Miso [11], and Zeebox [12]. They mostly deal with show details and social features, including TV program check-in, on-line chatting, and Facebook integration. IntoNow and Zeebox provide some kind of content related information, but they have no instant message and highly rely on specific content providers. RendezVous [13] aims to provide TV viewers interactivity and additional information about the program as well. It gathers and aggregates content related information from channel partners or their semi-editorial system. It is similar to our research. The main difference is we focus on automatic video and audio content analysis and event detection.

3 Design

In our prototype, 18 programs of various genres are selected, ranging from drama, variety show, talk show, sports, travel show, to movies. According to Goodman's research, people (i.e. website visitors) are particular interested in fan-based features and show annotations [14]. To better understand the audience of programs, we have interviewed potential viewers. Hundreds of ideas have been floated. Considering the limitations of time and resources, viewers' needs, innovation, and our technology capability, these ideas are boiled down to 6 categories (i.e. highlight, introduction, dynamic information, products, event, and statistics for sports only) of information that will be pushed to the second screen. Moreover, social features such as check-in, discussion forum and chatting are included in our prototype, too.

Fig.1 illustrates GUI samples of our prototype application. The "detection" button identifies the program currently playing on the TV, and then shows its highlight to present a brief introduction of the program to the viewers. In addition, viewers can easily access additional information through a feature menu as shown below.

(a) (b) (c)

Fig. 1. Second screen app UI: (a) main page (b) program highlight (c) dynamic information

1. Introduction. The section gives a brief introduction of the show such as the cast and summary. Viewers interested can further get more detailed information from related Wikipedia or Facebook via hyperlinks in this section. For certain genus (e.g. drama), a relation diagram among all characters is provided to help the audience follow the show more easily. For sports, the history and statistics of the players (or teams) is presented to the audience.

2. Dynamic Information. While the program is showing, the second screen app constantly and instantly shows the summary and updates the contents that range from plots, conversations, events, to products. This way, the viewers can well follow the program even if they have no time to watch the full show. In some countries (e.g. Taiwan), showing product placement is restricted though sometimes the audience would like to know about it. The second screen offers an alternative where the product's information can be pushed and displayed. Holmes [15] indicates that the second screen garners considerable visual attention, about 30% of the total viewing session. To minimize distraction to viewers, new-coming messages are placed at the top and old ones will be scrolled down and still kept on the UI. Information will be fed to the second screen during commercials to keep viewers glued.
3. Product. Items in the product placement are listed here. With a single click, TV viewers can mark whatever products they like as favorites, or even place an order. They can share the information of the product on Facebook. It serves as another way of promotion.
4. Event. The second screen can show instant messages about certain campaigns. Viewers can choose to participate and share the campaign information on Facebook.
5. Statistics. The statistics of a live Sports show is constantly and instantly updated, and made available to the viewers.
6. Check-in. Users share the information and comments of the program they are currently watching using Facebook. This very program will be further exposed to current viewers' social circles and promoted.
7. Discussion Forum/ Chatting Room. TV viewers communicate with one another using Discussion Forum for topic-independent messages and Chatting Room for free talk.

4 Prototype Implementation

A prototype based on the above design has been implemented. An Electronic Program Guide (EPG) containing 6 channels is used in our test, each representing a certain aforementioned genre. Three 5-minute long videos play repeatedly on each channel. A streaming server streams these programs, according to the EPG, and delivers related information to users as well. A push server is also installed.

4.1 Metadata Information Extraction

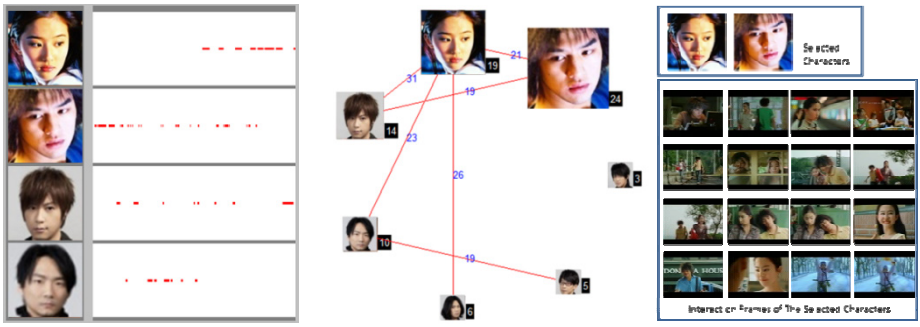
In our application scenario, we need to push additional information such as actors/actresses, products, location and events to the second screen. To effectively manage such information, the associated metadata from content owner is required, either generated manually or automatically. However, obtaining complete or sufficient metadata from TV companies is difficult, in practice. Hence we also develop content analysis technologies to automatically extract metadata from audio and video features. Initially, we focus on face recognition and tennis event detection.

With face recognition, we can identify the characters appearing in the show, and the relationship among them. Tennis event detection helps to detect events from a lengthy tennis game, and summarize it.

Face Recognition. Facial information is crucial that we can extract and manage/browse specific person in video contents. Nowadays accurate face recognition still poses a big challenge, although automatic face detection is mature and robust. To achieve robust face recognition, manual collection of faces as an identification basis is required. Our face recognition system consists of three steps which are face detection, face clustering, and recognition process.

First, a video is divided into scenes using frame-change detection. Lin and Liu's method [16] is used to detect human faces in each frame. Their detector is capable of detecting faces of different orientations. In a clip, close-by faces are clustered into a group. At the end, we apply face recognition to faces in each group.

After face recognition, the intervals of time when a character appears in the video are there. Generally speaking, the leading characters appear longer and more frequently in a video. Based on this assumption, we can compute the "importance" of each character, as shown in Fig. 2(a). The overlaps of the time intervals of two characters represent their correlation and degree of interaction. An interaction graph of characters can be constructed as shown in Fig. 2(b). Fig. 2(c) shows the interaction frames of the two selected characters based on the overlaps.



(a) (b) (c)

Fig. 2. Three applications based on human faces: (a) time-points of characters; (b) interaction graph of characters; and (c) interaction frames of the two selected characters

Tennis Event Detection. There are numerous video types, and oftentimes the video type classification is subjective. Each video type could need a content analyzer of its own, and each analyzer, itself, poses a big challenge to researchers and developers. We have implemented a content analyzer for tennis video event detection.

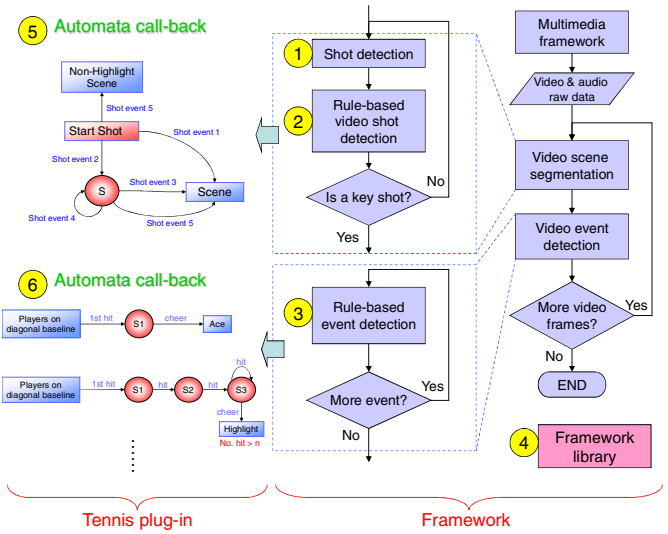


Fig. 3. Tennis video content analysis data flow

Below is brief explanation of the data flow Fig. 3. First, the content analyzer detects the shots, uninterrupted clips recorded by a single camera, by using a common video processing method. Then, rules are defined for scene segmentation and scene event detection. A scene is a collection of semantically related and temporally adjacent shots, depicting and conveying a high-level concept. In the scene event detection, some detectors compute the positions of the tennis court and the players. Please refer to [17] [18] for the details of the algorithm. We also have detectors to spot the sounds of hit and cheer. Most events occur during the rally between players, and are followed by cheers [17], if there are great shots. Audio plays an even more important role in event detection than video here. Audio features are cheaper to detect (than the video ones) and help significantly reduce the cost of the video processing (e.g. court and player tracking). Detected events will be exported in XML and pushed to viewer's second screen accordingly.

4.2 Audio Fingerprinting

Automatic synchronization between TV and second screen is a desirable feature for some TV viewers. The associated audio fingerprints are employed to detect the current TV program. We apply audio, over video, for easier management and computation. Moreover, recording audio is easier than recording video for most of the viewers. Push the "detect" button, and then app records the TV sound and instantly determines the program playing on TV.

An acoustic fingerprint is a condensed digital digest that deterministically computed from an audio signal. By comparing it with audio samples in an audio database [19], we can quickly find matched items. In our real time audio recognition system, the critical band analysis method from Kaller et al [20] is adopted to extract

and match fingerprints. A server computes the fingerprint of (1) an audio clip recorded by the client app and (2) a stream immediately after it is received from a channel, and updates the audio database.

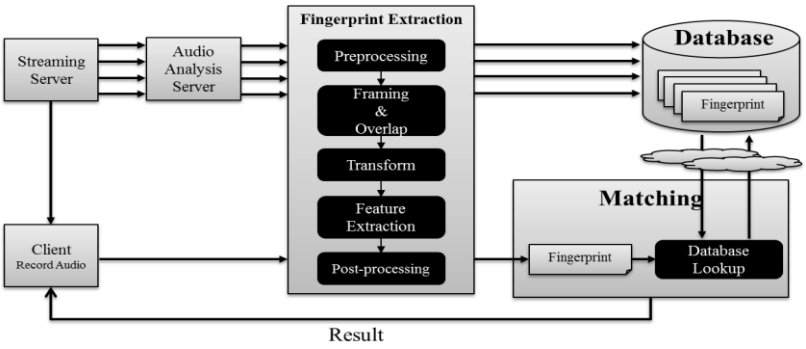


Fig. 4. Audio analysis system flow chart

5 Evaluation

We have conducted a usability test on our prototype. We have 20 participants from Taiwan (10 males and 10 females) of ages ranging from 22 to 40. Some of them are students, and others are our colleagues. None of them has ever participated in this project. They watch TV more than 1 hour per day on average, and use their mobile devices while watching TV at least once a week. From their feedback, we hope to better understand users' preferences and behaviors.

Since the second screen concept and apps are not popular in Taiwan yet, only 1 participant has ever used similar apps or services. However, 17 of them often search for additional information about TV on Google.

During the test, participants are asked to perform certain tasks in a simulated living room environment, which include (1) synchronizing TV with mobile device, (2) looking for additional information about a movie and the actress in the movie, (3) checking movie/ game status, and (4) purchasing a product. At the end, they each need to fill out a 5 Likert-based satisfaction questionnaire, and give comments on our prototype and second screen scenarios.

Overall, participants are positive about our prototype. They like the automatic program detection feature, and find it useful. Three of them would like our app to automatically detect channel change as well. The detection button on the app does not seem clear enough, and five participants suggest making it more eye-catching. They also show the concern about the reliability in a noisy environment. The information we provide in our app is appealing to participants. They particularly like the program introduction, dynamic information, products, and statistics. All participants would like to use this app in the future if it is free. If they have to pay, then more incentives are needed.

As for the usability of our prototype, participants are able to complete all assigned tasks, except for reviewing missing TV contents. Based on our test scenario,

participants turn on TV to watch the movies that have already been playing for some time. To follow the movie, they are asked to check the content of the part they miss. Thirteen participants mistakenly click "introduction" instead of "dynamic information". They assume that TV content summaries should be placed in the program introduction session. Moreover, four participants click "introduction" to check personal statistics in a game. From their behaviors, we conclude that people tend to click "introduction" when they are lost. Therefore, the wording of the menu needs to be clearer and more intuitive. Ten participants click the wrong icon when synchronizing second screen with TV program. It is because the "Detect program" icon is too small to be distinguished from its logo.

When asked about the wish list for new features, six would like program recommendation, and four would like to control their TV directly with the app. TV schedule and more program related information (such as movie review, behind the scenes, game strategy) are desirable.

6 Conclusions and Future Work

In this paper, we propose second screen scenarios, which focus on the real-time delivery of additional program information. Our goal is to enhance the TV viewing experience with a second screen, and such information delivery should not interfere with users' normal TV viewing behavior, and additionally provide new business opportunities. An Android-based prototype app has been implemented. This app automatically detects TV program, and instantly pushes related information, such as program introduction, dynamic information, product, and statistics. Audio fingerprinting, face recognition, and tennis event detection technologies are employed in this app.

To verify our development work, an experiment involving 20 participants shows that people adopt the second screen quickly, and like it especially when this application is free. However, participants would like more diverse features and richer information on the second screen.

In the future, we plan to expand the scalability of our prototype to support more concurrent connections. To enrich our contents, we will continue developing new media analysis techniques. For example, live audio/video keyword detection and spotting is on the drawing board. We will iteratively revise the design of our system according to the feedback from users, and hope to release our product to the market as soon as possible.

References

1. Red Bee Reports. Broadcast industry not capitalising on rise of the second screen: New research reveals that majority of TV viewers now dual screen but industry's take up of synchronous apps remains slow (2012), http://www.redbeemedia.com/sites/all/files/downloads/second_screen_research.pdf
2. Nielsen Consumer Reports. State of media: consumer usage report (2011), http://blog.nielsen.com/nielsenwire/mediauniverse/

3. Tsekleves, E., Whitham, R., Kondo, K., Hill, A.: Investigating media use and the television user experience in the home. Entertainment Computing 2(3), 151–161 (2011)
4. Cruickshank, L., Tsekleves, E., Whitham, R.: HILL, A, and Kondo, K. Making interactive TV easier to use: interface design for a second screen approach. The Design Journal 10(3) (2007)
5. Fallahkhair, S., Pemberton, L., Masthoff, J.: A dual device scenario for informal language learning: interactive television meets the mobile phone. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies, pp. 16–20 (2004)
6. Fallahkhair, S., Pemberton, L., Griffiths, R.: Dual device user interface design for ubiquitous language learning: mobile phone and interactive television (iTV). In: Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE 2005), pp. 85–92 (2005)
7. Basapur, S., Harboe, G., Mandalia, H., Novak, A., Vuong, V., Metcalf, C.: Field trial of a dual device user experience for iTV. In: Proceedings of the 9th International Interactive Conference on Interactive Television, pp. 127–136 (2011)
8. Basapur, S., Mandalia, H., Chaysinh, S., Lee, Y., Venkitaraman, N., Metcalf, C.: Proceedings of the 10th European Conference on Interactive TV and Video, pp. 87–96 (2012)
9. IntoNow Application, http://www.intonow.com/ci
10. GetGlue Application, http://getglue.com/
11. MISO Application, http://gomiso.com/
12. Zeebox Application, http://zeebox.com/tv/home
13. Poivre, S.R.: An editorial ruled based contextual TV information system. In: Adjunct Proceedings of the EUROITV 2011, p. 96 (2012)
14. Goodman, J.M.: Enhanced TV features on national broadcast and cable program web sites: An exploratory analysis of what features are present and how viewers respond to them. Master's Thesis, Scripps College of Communications of Ohio University (2009)
15. Holmes, M.E., Josephson, S., Carney, R.E.: Visual attention to television programs with a second-screen application. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 397–400 (2012)
16. Lin, Y.Y., Liu, T.L.: Robust Face Detection with Multi-class Boosting. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 680–687 (2005)
17. Chu, W.T., Tien, M.C., Wang, Y.T., Chou, C.W., Hsieh, K.Y., Wu, J.L.: Event detection in tennis matches based on real-world audiovisual cues. In: Proceedings of the 20th IPPR Conference on Computer Vision, Graphics, and Image Processing, pp. 541–548 (2007)
18. Tien, M.C., Wang, Y.T., Chou, C.W., Hsieh, K.Y., Chu, W.T., Wu, J.L.: Event detection in tennis matches based on video data mining. In: Proceedings of IEEE International Conference on Multimedia & Expo, pp. 1477–1480 (2008)
19. Acoustic fingerprint – Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Acoustic_fingerprint#cite_note-1
20. Haitsma, J., Kalker, T.: A Highly Robust Audio Fingerprinting System. In: Proceedings of the 3rd International Conference on Music Information Retrieval, pp. 107–115 (2002)