CircleBuy: A Visual Search Based Second Screen Application of Buying Products in Videos

Yucheng Jin
KU Leuven
Celestijnenlaan 200A
B-3001 Leuven, Belgium
yucheng.jin@cs.kuleuven.be

Joris Klerkx
KU Leuven
Celestijnenlaan 200A
B-3001 Leuven, Belgium
joris.klerkx@cs.kuleuven.be

Katrien Verbert
KU Leuven
Celestijnenlaan 200A
B-3001 Leuven, Belgium
katrien.verbert@cs.kuleuven.be

ABSTRACT

Visual search aims to find information based on visual features of objects, and it is extremely useful when specific keywords for describing an object are unknown. Visual search has been considered as one of the most exciting trends in e-commerce. *CircleBuy* is a second screen application that leverages visual search to allow users to select pieces of furniture in a video and to receive similar product recommendations. In addition, our application attempts to support users in making purchase decisions by showing a 3D model of the piece of furniture that helps them to explore the product in detail. Products can furthermore be shared on television to discuss them with friends and family. This paper presents the concepts behind *CircleBuy*, its engineering and implementation aspects, and its evaluation. The concept was evaluated with 22 subjects in a simulated home environment. The results demonstrate the meaningfulness of the concept and offer insight into the utility and usability of the design of such a visual search based application.

ACM Classification Keywords

H.5.2 User Interfaces: Screen design

Author Keywords

Visual search; second screen; online shopping; user experience.

INTRODUCTION

Mobile devices have become indispensable in our daily life. Increasingly, viewers use mobile devices while watching television [13]. When the user for instance sees an appealing product in a video, she can search for this product on her mobile device. However, this search activity may break the experience of watching videos, and it is tedious to come up with effective keywords for searching this product. Moreover, advertisements in a video are often minimally linked to the content of the video itself [9]. These two concerns lead us to employ visual search technology to facilitate the process of searching products in videos. The strength of visual search is the ability to identify visual features of objects such as colour, shape, and size, which facilitates the product search as it obeys the web usability adage "do not make me think" [3].

We adopted a design based research approach to investigate if the concept of using visual search to explore similar products in videos is meaningful. For this, *CircleBuy* was developed that empowers users to select furniture and decor items of their interest at any time in the video. The application will then automatically issue queries to online shopping websites for searching similar products. Those results are then presented back to the user. Furthermore, the application has a view showing 3D models of products and a function of sharing products on television.

To evaluate its meaningfulness, we conducted a user study (N=22), in which subjects were asked to watch four videos and select the furniture and decor items that interested them in videos. We asked the subjects to give feedback via a questionnaire. In addition, we recorded the subjects use of application and collected the subject thoughts using think aloud protocol. Preliminary evaluation results indicate that visual search technology can facilitate the process of buying furniture that appears in videos. The more similar the results are compared with the objects in the videos, the more satisfied users tend to be. The representation of 3D models and the sharing feature add limited value to making purchase decisions.

The remainder of this paper is organised as follows: we start with an overview of related work. Then we describe the design and implementation of our application. Results of the user evaluation are presented next. We conclude with a discussion and suggestions for future work.

RELATED WORK

Several researchers have investigated the use of second screen for interactive television. Cesar et al. [4] present an architecture of utilising second screens to enhance user experience of interactive television, which supports controlling, enriching, and sharing of content. Simon et al. [10] employ mobile devices as second screens to synchronise the content that is displayed on television and to share experiences on social media. Thus, second screen can be seen as an effective tool to enhance interactivity of television. Moreover, a field trial shows enjoyable and immersive user experience of using a

EICS'16, June 21-24, 2016, Brussels, Belgium
ACM 978-1-4503-4322-0/16/06.
http://dx.doi.org/10.1145/2933242.2935871

dual device while watching television [1]. Based on this state-of-art second screen design for interactive television, we followed certain guidelines to design our application.

Several other researchers have presented various approaches to search for objects in videos using visual search technology. Sivic et al. [11] describe an approach to rapidly and accurately localise all occurrences of an object in videos, given a query image of object. Girod et al. [5] present a mobile visual search based application and its architecture optimising the process on both mobile clients and server. Yeh et al. [15] propose a mobile image-based search system that takes images of objects as queries and finds relevant web pages containing similar images. Furthermore, users can specify objects of interests via an interactive segmentation tool. Our application offers an integrated solution to specify and search for objects of interest in videos.

In addition, a number of mobile apps of visual search such as Slyce [12], Amazon Flow [6], Snap Fashion [8] have been developed, in which users can find products by taking a photo. However, taking a photo to precisely select a specific product in a video is difficult. *CircleBuy* is designed especially for online shopping of pieces of furniture and decor with seamlessly integrates visual search while watching videos. There also exists a second screen application named TheTake [14] which allows users to discover products from their favourite videos. However, all videos in TheTake have been processed beforehand and annotated manually. While this approach ensures high quality of recommend products, it does not scale well compared to the abundance of existing products in videos. On the contrary, our work attempts to leverage state-of-art visual search technology to allow exploration of all pieces of furniture that appear in the video. Although the quality of recommended products in this approach is admittedly lower, we want to investigate if it is feasible and meaningful approach.

SYSTEM DESIGN AND IMPLEMENTATION

CircleBuy is a second screen application that employs visual search technology to support users in searching and buying furniture and decor in videos. We describe general functionality, system architecture, and finally the user interaction design.

Functionality Design

The second screen application *CircleBuy* allows users to control the television by selecting videos to play, pause, rewind and fast-forward. Users can furthermore search for pieces of furniture that appear in those videos. Selecting a piece of interest is done by drawing a circle around them (i.e elliptical selection). This selection will result in a list of similar products offers in online web shops. To aid users in the process of deciding which one of these products to buy (or not), *CircleBuy* provides a 3D view that enables users to explore more details of the product. To ask a second opinion from the other users in the room, users can show the selected piece on the television.

System Architecture

Figure 1 shows the system architecture of our demonstrator. The system consists of two separate frond-end applications

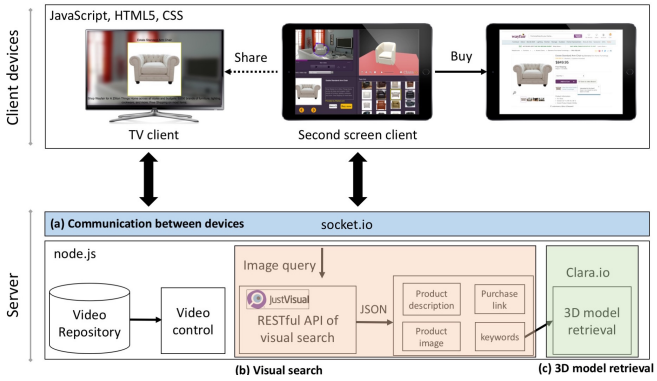

Figure 1. System architecture of demonstrator.

that communicate with each other: a TV that plays a video, and the second screen app *CircleBuy* that runs on a tablet and is synchronised with TV. We implemented *CircleBuy* based on above design. Both web applications were built using HTML 5, JavaScript and CSS.

The back-end architecture is implemented with node.js, and consists of three main services: (a) a communication service to synchronise TV and second screen, (b) a visual search service, and (c) a 3D model retrieval service. In addition, the server contains a video repository and a module of video control. The video repository contains four genres of videos, each of which has four 4-5 minutes long videos. The server streams these videos and delivers searched product information to users as well.

(a) Communication between devices

The real-time communications between the second screen and the TV was implemented by socket IO which is a JavaScript library for real-time web applications. A node.js web server application handles the various socket.io requests such as pause video and sharing a product on TV. In particular, the server first identifies an event after receiving a request, and then it decides what should be responded to all clients listening to the event.

(b) Visual search

Visual search has been implemented by using the JustVisual API for furniture [7] which can match any image to over 10 million furniture and decor products. To allow the user to select an object from a video, a canvas of HTML5 is created as a transparent overlay on top of the video player when the video is paused. The selected area is sent as an image query. After matching the searched image with the indexed product images in the product repository, the information of resulting products is returned in a JSON file which contains product description, product image, purchase link, manufacturer of product, and tags of the recognised object.

(c) 3D model retrieval

3D model retrieval relies on keywords matching: the keywords of visual search results are matching to the tags of

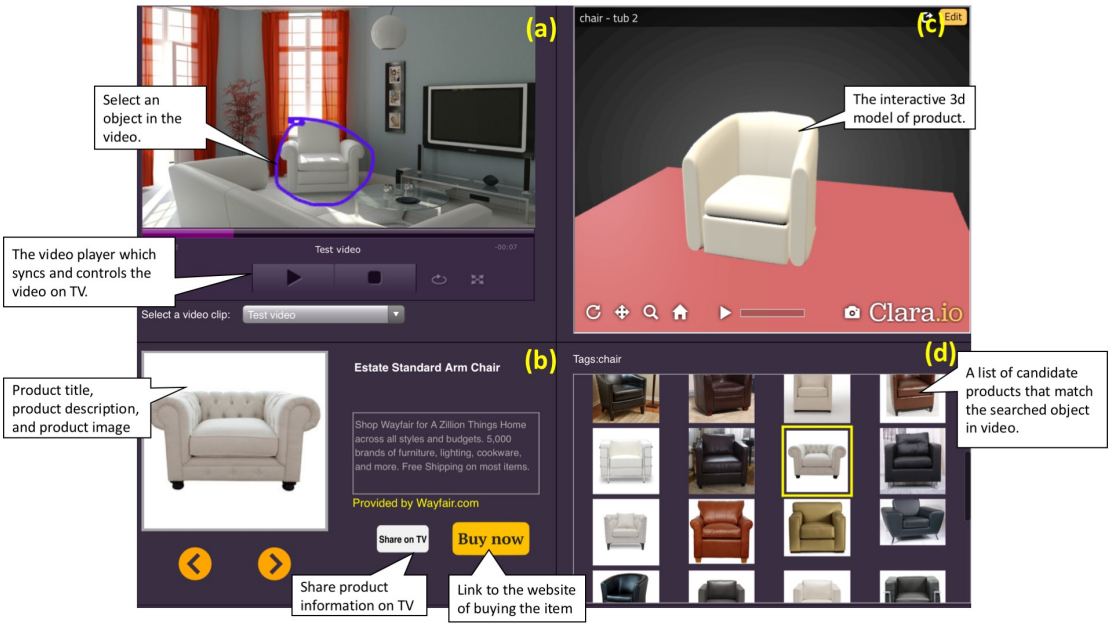

Figure 2. User interface of *CircleBuy*: (a) video player with playback controls and a transparent overlay for drawing a circle around an object in the video; (b) product information window; (c) 3d view of product; (d) a list of candidate products.

the 3D model. We employed Clara.io, a free web-based 3D modelling software to create, tag, and render the 3D model, and then we used iframe to embed the 3D view in our web application. The 3D model retrieval is implemented with a naive keywords matching rule. Most tags correspond to categories of furniture such as chair, sofa, and bed. Once a tag of the 3D model hits the "keywords" of visual search results, this 3D model is selected.

User Interaction Design

Figure 2 shows an overview of the user interface of *CircleBuy*. The user interface consists of four connected parts: (a) video player, (b) product information, (c) 3D view of product, and (d) candidate resulting products.

In Figure 2.a, the user can select a video from the drop-down list to watch. The video will be updated on the TV when a new video is selected on the second screen. The video player provides controls such as pause and fast forward. In order to facilitate object selection in videos, an elliptical selection tool is implemented. After selecting a piece of furniture, product information of a similar product is shown in Figure 2.b, including title, description, image, merchant, and a purchase link of the product. The arrow buttons allow the user to explore other candidate products. The button "Share on TV" allows the user to share her favourable product on TV for discussion with friends or family. While performing this interaction, the video will be paused and a blur effect will be added to the video as shown in Figure 3. The button "Buy now" leads the user to a web shop selling the selected product. In addition, in order to support users to explore the product in detail, the application shows a 3D view of the product (see Figure 2.c), where the user can zoom in/out and rotate the 3D model. More images of

Figure 3. The scenario of sharing a product on TV.

candidate products are shown in a list (see Figure 2.d), which offers a quick overview of resulting products. Thus, the user can judge if the quality of the visual search is satisfactory and otherwise make a new selection. When a new product is selected from the list of candidate products, it will be labelled with a yellow rectangle and the content of product information window and 3D view will be updated accordingly.

EVALUATION

Method

We employed a mix of quantitative and qualitative methods to evaluate if our approach can be meaningful towards facilitating the process of purchasing pieces of furniture that appear in videos. More specifically, we investigated if:

R1 : *CircleBuy* facilitates the entire process of finding and buying products that appear in videos.

Figure 4. Evaluation setup.

R2 : The similarity of resulting products of visual search influences users' satisfaction and therefore user adoption.

R3 : Showing a 3D model view adds value for making purchase decision(s).

R4 : Sharing a product on TV adds value for making purchase decision(s).

We first asked subjects to fill out a pre-study questionnaire before evaluation. Afterwards, we introduced the setup of the evaluation and provided a short tutorial for using the application: i.e. we played a test video on TV and tablet and showed users how to search for a product from a video and how to share the product on TV. Once the subjects indicated that they were familiar with the application, they were asked to spend at least ten minutes watching a video from each of the four video genres and search for furniture and decor items of interests in videos. During the evaluation we also employed the think-aloud protocol to collect subjects' feedback. In the end, subjects were asked to fill out a post-study questionnaire. All statements of the questionnaire were rated by using a 5-point Likert scale, ranging from Strongly disagree to Strongly agree.

Apparatus

We set up our evaluation in an office with a 60 inch TV and a sofa (see Figure 4) to mimic a living room setting. The video played on the TV was controlled by a Mac mini. The second screen application was running on an iPad Air 2. The whole procedure of evaluation was recorded by a Logitech camera.

Demographics

To get initial feedback about *CircleBuy*, we evaluated the application with 22 subjects (see Table 1 for demographics). The subjects were compensated with snacks and drinks. Prior to the evaluation we interviewed subjects with some questions about their experience of searching products in videos. 68% subjects used keywords to search for products in Google and the rest of the subjects just visited e-commerce websites such as Amazon and Taobao directly. The large majority of subjects indicated that it is tedious to use keywords to search products (median = 2) and that it is not easy to search products from video by using keywords (median = 2).

RESULTS

In this section, we will report the quantitative results, questionnaire results and subjective feedback of our evaluation.

22 subjects, age: 24-37 (Mean = 29, SD = 3.86)	**%**	**N**
male subjects	63.6	14
searched for products in videos	86.4	19
used second screen while watching videos	27.3	6
used visual search technology	50	11

Table 1. Demographics of subjects.

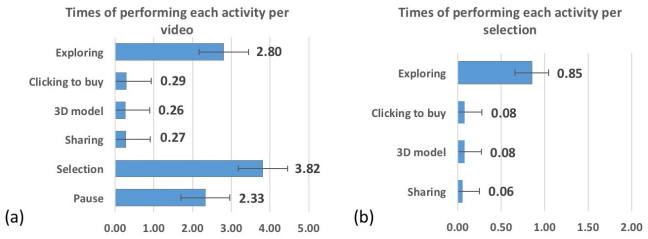

Figure 5. (a) times of performing each activity per video; (b) times of performing each activity per selection.

Quantitative Results

Usability test

We firstly employed the System Usability Scale (SUS) [2] method to test the usability of our application. The average SUS score is 80.8 and the median is 82.5. According to the SUS score explanation, our system gets an A score (the top 10% of scores), which means overall the application has an excellent usability [2]. This indicates that usability issues minimally influenced subjects' opinions on meaningfulness.

Log data results

We recorded all subject activities when they were using our application. In particular, we calculated the times of pausing a video, the time of searching a product from the video, and the times of interacting with other components of application such as the 'Share on TV" button, 3D model view, "Buy now" button, and the list of candidate products.

Figure 5.a shows that the most frequently performed activities are "select objects in videos" (mean: 3.82; SD: 2.10),"pause vide" (mean: 2.33; SD: 1.52), and "explore other candidate products" (mean: 2.80; SD: 1.70). While the activities of"sharing" (mean: 0.27; SD: 0.59), "interacting with 3d models" (mean: 0.26; SD: 0.54) and "clicking to buy" (mean: 0.29; SD: 0.56) were performed much less frequently.

In addition, Figure 5.b reveals that the subjects tend to more likely explore other candidate products (mean: 0.85; SD: 0.54) after searching a product from a video. Whereas, the possibility of performing other activities are relative low, "sharing" (mean: 0.06; SD: 0.12), "interacting with 3d models" (mean: 0.08; SD: 0.19) and "clicking to buy" (mean: 0.08; SD: 0.17).

Questionnaire Results

Control and search

Most of the subjects tend to control the video play on the second screen (median = 4). Compared with the traditional approach used for searching products in videosn such as a keywords search in Google, the responses of subjects indicate that

our application makes the process of searching more effective (median = 5) and easier (median = 5). Furthermore, subjects have more fun while using our application (median = 4.5). Almost all subjects would like to use the elliptical selection tool to select an item on the video player (median =5); and most of subjects also thought the visual search is integrated with video watching properly (median = 4).

Making Purchase Decisions

As for the support in making purchase decisions, it seems that the provided product information is not sufficient to make purchase decisions (median = 3). Moreover, subjects perceived the usefulness of 3D models (median = 3.5) and sharing (median = 3) to be limited. However, it is more likely that subjects can find a favourable product (median = 4) from the list of candidate products. Subjects also indicated that they would like to click "buy" button to check more purchase information (median = 4) and share it on TV (median = 4).

Subjective Feedback

General opinions

In general, the large majority (N=19, 86%) of subjects gave positive feedback to the concept of the application. They found the feature of visual search for products in a video to be a welcome addition while watching videos. Around two third of subjects (N=14, 64%) responded that they would like to use such an application if they intend to look for some appealing objects in videos. Two subjects gave negative feedback. One was reticent to engage with the multiple features of the application, describing a desire to just watch TV without interference from the tablet. In the following sections, we describe specific behaviour and opinions of subjects central to some of our research questions.

Video control

The sync function is not always necessary. Sometimes it makes object selection difficult. More than a quarter of subjects (N=6, 27%) stated that the time of switching attention from television to second screen make them fail to pause the video at a right time thus they have to rewind the video.

Visual search result

The degree of match tends to affect users' satisfaction. Users expect to see the perfect match of products they selected. More than two thirds of the subjects (N=15, 68%) only want to see products of perfect match. This includes having the same style and colour. Less than one third of the subjects are happy to also see products of rough match.

3D models of products

It is controversial to show the 3D models of products, especially when they are not aligned with the 2D images of product. Around two thirds of subjects (N=14, 64%) would not like to check the 3D models of the product. S22 thought the product specification and multi-view images allow him to know the product quite well. S1 suggested to show the 3D model via an AR view.

Sharing

The usefulness of sharing depends on user preference and context of use. We found more than two thirds of subjects (N=15, 68%) tend to search and check products without sharing. S9 considered the sharing as an obtrusive behaviour for watching television. S17 would like to just pass the second screen device to other people if she wants to get their feedback. However, S3 would like to share the product on TV if his family is not sitting close to him.

Purchase information

Price preview and consumer review are crucial to making purchase decisions. However, due to the limitations of the API we cannot obtain this information in our application. More than half of subjects (N=12, 55%) thought our application should provide additional product information such as price and reviews. The subjects became frustrated when they found a product that is too expensive on the shopping website.

DISCUSSION

The responses to the questionnaire have shown that *our app **can** facilitate the entire process of finding and buying products that appear in videos* (R1). The finding corresponds to the observations during the think-aloud study, where most subjects were able to use the application very quickly after a short tutorial. Despite occasional mismatches, the majority of subjects still perceived our application as a more effective and interesting way to search for products in videos. Moreover, the high enjoyment of using the application may increase user engagement.

The results of the questionnaire together with the subjective feedback show that the majority of subjects prefer to see the results of perfect match rather than the results of close match. This is reasonable since normally users expect to find the exactly same product when they are searching for products in videos. Therefore, we can assume that users tend to be more satisfied with the highly matched products especially in terms of style and colour. However, some subjects perceived the products of close match or even mismatch as an increase of diversity of search results. Therefore, *the similarity of resulting products of visual search **tends to** influences users' satisfaction and therefore user adoption* (R2).

According to the quantitative results and the responses to the questionnaire, the subjects appear to be more neutral to the 3D models and product sharing in terms of support in making purchase decision. The log data indicates that the number of "interacting with 3d models" and "sharing products on TV" activities are quite low. As S4 mentioned, it is more likely that a user decides to buy a product based on its price and consumer reviews. In addition, the usefulness of the 3D model is also limited due to its imperfect match with the product and the longer rendering time. Thus, our current support for *showing a 3D model view adds **limited** value for making purchase decision(s)* (R3). The usefulness of sharing varies in different contexts of use and it is also affected by user preference. Therefore, it is more likely that *making purchase decision(s) has **certain prerequisites** to benefit from sharing a product on TV* (R4).

CONCLUSION

We have presented a visual search based second screen application, by which the user can search and buy products in

videos more effectively and easily. The results of our evaluation provide initial insight and guidance for practitioners and developers for building such a visual based application. We are planning to show a demonstration of the *CircleBuy* at the conference on a tablet as well as on a projector (main screen). We believe our demonstration can bring some new ideas that can stimulate a discussion on building visual search based shopping applications.

Future work is needed to improve the synchronisation of second screen and TV and the quality of 3D models.

ACKNOWLEDGEMENTS

We would like to thank the reviewers for their helpful feedback and suggestions. Part of this work has been supported by the IWT Flanders, grant agreement no. IWT-SBO-Nr.110067 (PARIS).

REFERENCES

1. Santosh Basapur, Gunnar Harboe, Hiren Mandalia, Ashley Novak, Van Vuong, and Crysta Metcalf. 2011. Field trial of a dual device user experience for iTV. In *Proceddings of the 9th international interactive conference on Interactive television*. ACM, 127–136.
2. John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
3. Linda Bustos. 2014. Ecommerce Trend: Keep Your Eyes on Visual Search. (2014). Retrieved April 10, 2016 from `http://www.getelastic.com/ecommerce-trend-keep-your-eyes-on-visual-search/`.
4. Pablo Cesar, Dick CA Bulterman, and Jack Jansen. 2009. Leveraging user impact: an architecture for secondary screens usage in interactive television. *Multimedia systems* 15, 3 (2009), 127–142.
5. Bernd Girod, Vijay Chandrasekhar, Radek Grzeszczuk, and Yuriy A Reznik. 2011. Mobile visual search: Architectures, technologies, and the emerging MPEG standard. *MultiMedia, IEEE* 18, 3 (2011), 86–94.
6. A9.com INC. 2015. Flow Powered by Amazon: Discover the world around you. `http://www.a9.com/whatwedo/mobile-technology/flow-powered-by-amazon/`. (2015). Accessed: 2016-03-05.
7. JustVisual. 2016. JustVisual API. `http://justvisual.com/furniture-api/`. (2016). Accessed: 2016-04-01.
8. Snap Fashion Ltd. 2016. Shop fashion in a Snap. `https://www.snapfashion.co.uk`. (2016). Accessed: 2016-03-05.
9. Tao Mei, Xian-Sheng Hua, Linjun Yang, and Shipeng Li. 2007. VideoSense: towards effective online video advertising. In *Proceedings of the 15th international conference on Multimedia*. ACM, 1075–1084.
10. Heloisa Simon, Eros Comunello, and Aldo Von Wangenheim. 2013. Enrichment of interactive digital TV using second screen. *International Journal of Computer Applications* 64, 22 (2013).
11. Josef Sivic and Andrew Zisserman. 2009. Efficient visual search of videos cast as text retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31, 4 (2009), 591–606.
12. Inc Slice Technologies. 2016. Slyce: The Complete Visual Search Solution. `http://slyce.it/`. (2016). Accessed: 2016-03-05.
13. Aaron Smith and Jan Lauren Boyles. 2012. The rise of the "connected viewer". *Pew Internet & American Life Project* (2012).
14. TheTake. 2016. TheTake. `https://thetake.com/`. (2016). Accessed: 2016-04-01.
15. Tom Yeh, Kristen Grauman, Konrad Tollmar, and Trevor Darrell. 2005. A picture is worth a thousand keywords: image-based object search on a mobile platform. In *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 2025–2028.