Towards a Multi-Screen Interactive Ad Delivery Platform

Francisco Vega*, José Medina†, Víctor Saquicela*, Kenneth Palacio-Baus†, and Mauricio Espinoza*
*Computer Science Department, University of Cuenca, Ecuador
{francisco.vegaz, victor.saquicela, mauricio.espinoza}@ucuenca.edu.ec
†Department of Electrical, Electronic Engineering and Telecommunications, University of Cuenca, Ecuador
{jose.medina, kenneth.palacio}@ucuenca.edu.ec

***Abstract*—Interactive advertising based on multiple devices opens new possibilities for mobile applications, where users can search, select, or expand the information provided in advertising commercials by incorporating interactivity-friendly companion devices such as smart phones and tablets. In this paper, we derive the basic requirements for a flexible infrastructure that can support interactive ad applications. The infrastructure comprises a set of components and their externally visible properties, and the relationships between them. The main contributions of this work are first, establishing a desirable set of requirements for a suitable working scenario on which the different interactivity supporting systems people use in a regular basis can be easily integrated, and second, the definition of a platform which considers the different stages and requirements identified from different works in related areas.**

***Index Terms*—interactive advertising, ad delivery, multi-screen**

I. Introduction

Interactive advertising is gradually replacing traditional linear advertising, improving how brands and broadcasters reach customers. This emergent communication channel has reached digital platforms such as television and radio, which are traditionally mass and one–way media. Multiple reports [1], [2], [3] show that new generation users of radio and television are not satisfied with a single screen experience and expect content to be seamlessly delivered across multiple devices and screen types instead.

From the viewpoint of the digital advertisement service provider business, this trend offers new opportunities and challenges. On one hand, a multi–screen environment which incorporates users into new advertising experiences that can be exploited by advertisers. On the other hand, changes in consumer behavior require ad delivery platforms with new technical complexities, which have not previously faced by the media industry.

Until now, the most broadcasters have been building dedicated second screen apps to complement their programming and to capture the attention of users during an ad break. However, this solution limits the audience and therefore the potential for second screen synchronization and personalization in the advertising industry. According to different companies such as Mufin[1], Civolution[2], or WPP[3], which are dedicated to identifying, managing and monetizing audio and video media content, what the advertising industry needs is a more holistic approach, where be possible to delivery synchronized and personalized ad content on all consumer's devices (smart phones, tablet, personal computer, work computer, etc.), then, the audience segments could be expanded to include all the relevant devices.

In our opinion, the inclusion of cross-device user identification data is a mandatory requirement for the creation of multi-screen interactive ad delivery systems, because this brings a double advantage to advertisers: i) the possibility of reaching users in a more efficient way, for example, by sharing the user known data on a device to his profile on another device, and ii) the possibility of increasing the user experience by having a personalized access to the content, regardless of used device.

Existing academic projects have paid more attention to solve the challenges of Internet advertising, which is one of the most commonly used marketing channels. In the context of multimedia advertising, the most effort has been put on targeting advertising. To the best of our knowledge, very few projects have attempted to improve the efficiency of both digital video and audio advertisements. In [4] the authors combine online shopping information with an advertising video to direct viewers to proper online shopping places. The work presented in [5] takes advantage of an internal reasoning about items and user profiles to automatically compose interactive applications offering personalized commercial functionalities on digital television (DTV). While these works offer important technical details to improve interactivity in digital multimedia ads, none of these works approach the challenges of the adoption of a multi-screen advertising strategy nor cover advertiser requirements such as understanding the perceptions and usage patterns of multi–screens users. Moreover, current commercial systems such as LiveSync[4], Beamly[5], or AdsWizz[6], suffer

978-1-5386-3057-0/17/$31.00 © 2017 IEEE

[1]https://www.mufin.com/
[2]http://www.civolution.com/
[3]http://www.wpp.com/wpp/
[4]http://wywy.com/products/livesync/
[5]https://www.beamly.com/
[6]http://www.adswizz.com/products/

from the lack of a precise and clear functional requirements and a comprehensive architecture.

In this paper, we address the challenges for enabling interactive digital video and audio ads (henceforth digital ads) through second–screen devices. Our contributions are: first, we describe some desirable requirements for an interactive ad delivery platform motivated by works in related areas and own experiences; and second, we outline our view of the definition of a platform architecture that considers requirements identified from different related works.

The remainder of this paper is organized as follows. In Section II we offer a definition of interactive advertising as it is perceived in this work. Section III illustrates with the aid of a specific scenario the operation of a platform to deliver interactive ads on different companion devices. Section IV describes some desirable requirements for an interactive ad delivery platform. In Section V we introduce a general architecture to support this kind of platforms. The demonstration of the platform functionality through two prototype applications is described in Section VI. Conclusions and future work are presented in Section VII.

II. Definition of Terms

According to the editors of the Journal of Interactive Advertising[7], this term is defined as "paid and unpaid presentation and promotion of products, services and ideas by an identified sponsor through mediated means involving mutual action between consumers and producers" [6].

Interactive advertising as we perceived it in this work describes the fact that the promotion of products, services and ideas is a two-way conversation and not just the one-way communication between consumers and producers. We believe that remaining terminology, used to describe our platform, cannot be fully or correctly understood without being contextualized in reference to a number of underlying trends that have enriched the development of interactive advertising. These trends (inspired in the work of [7]) are:

- *The revolution in technology*. Literature research [8], [9] has revealed that the management of effective digital video and audio marketing campaigns at scale cannot be achieved without the use of i) a companion device (e.g., tablet, mobile phone, etc.) –here multi-screen– that support the Internet browsing and multimedia content delivery to permit advanced consumer feedback, offering rich communication capabilities for advertisers, and ii) semantic technology that enables to marketing stakeholder describing their documents and requirements in an intuitive and flexible manner, so that computers to know what particular documents mean, and to understand the relationships between them.
- *The revolution in marketing thought*. As described in [7], there has been a clear evolution from mass communication to two-way communication to interactive forms of communication. This new trend involve the need of understanding of the customer through of different behavior targeting techniques, where the inferred information on a user's behavior serves to understand his or her needs and deliver personalized advertisements accordingly. Again, the semantic understanding of the data may allow to shift from the process of marketing management from the firms' point of view to a "customer-centric" focus.
- *The revolution in communication and distribution channels*. We agree with [10] in that the properties of pervasive computing (automation, interactivity, ubiquity) make it a powerful tool for achieving the goal of advertising; that is, to impart information, evoke emotions, and trigger actions. Also, pervasive computing technologies as digital signage and mobiles sensors may help to offer localized and context-aware advertisements to potential customers [11]. Regarding distribution channels, now have more ways to reach the customer than ever, including various digital video and audio platforms. To reach the full potential on these advertising channels is very important because this market spent only in the USA approximately $187 billion, of which 42% (nearly $79 billion) belonged to advertisement for television and 10% (17,6 billion) for radio[8]. Moreover, different reports shown the effectiveness of these platforms regarding traditional advertising, because combines the strength of conventional advertising in TV, radio and the internet [12], [13], [14].

The next section describes a possible scenario of an ad delivery platform.

III. Scenario

To illustrate how the proposed platform could be used to deliver personalized interactive advertisements on second screen devices, consider the following scenario:

Let us imagine a user living in United States (US), who is watching his favorite television program. Then, the television program is interrupted for a few moments to give way to ad break. During the advertising space, one of the commercials captures the user's attention, in particular the advertising about the launch of the new vehicle Nissan Rogue. At this moment, the user giving a click on his mobile device is synchronized with the advertising of his interest. In few seconds the application detects the commercial that is being transmitted in live. Furthermore, the application automatically recollects additional information such as: the geographic location where the user is connected (e.g., US), the advertising type (e.g., an advertising about vehicles), the ad scope (e.g., if the advertising is valid only for the US), or the type of user's companion device used for the connection (e.g., smart phone or tablet).

Once the information is collected, a server sends to the second screen application multiple data links that allowing for

[7]http://jiad.org/

[8]Report obtained from Strategy Analytics (https://www.strategyanalytics.com).

instance, to obtain additional information about vehicle model or the nearest dealers where is possible to obtain discounts or special promotions. In addition, let us imagine that at this point the user decides to switch his companion device to a tablet instead of his smart phone. Again, the application detects the type of device and it gives to the user additional information to the one deployed in the smart phone. An example of this additional information could be a data link that allows the user to build a custom vehicle version, by selecting different features such as: the exterior color, the interior color, or the possibility of incorporating additional accessories as roof rail crossbars or splash guards.

Additionally, let us suppose that a user outside of the US has access to the same commercial via cable or satellite television. Once the contextual information such as the user´s geographic position or the data that identifies the commercial has been collected, the platform determines that in the country where the user is watching the commercial, the Nissan Rogue model is not distributed, but instead the vehicle is called Nissan Xtrail. In this case, the additional information to be displayed on the companion device should correspond to information on this vehicle model.

Finally, to complete the scenario explanation, let us suppose that a third user is interested in the advertising described previously and let us imagine besides that in the country where the user lives, the brand Nissan is not marketed. In this case, the application should show information about similar vehicle models belonging to other brands in the companion device. To do this, the platform must be able to automatically recognize that the Nissan Rogue model belongs to the group of Crossover or Sport Utility Vehicle (SUV) vehicles.

IV. Platform Requirements

In this section we present some desirable requirements for the platforms that manage interactive advertising on digital audio and video channels using companion devices, motivated by works in related areas and own experiences. To define infrastructure requirements, we have been collecting key factors from different solutions, comparing them with our own observations in the field of interactive advertising, and grouping them according to their nature and relationship into four groups: i) cross-device synchronization; ii) content personalization; iii) content authentication; and iv) scalability.

- *Cross-device synchronization.* The use of multiple devices is a recent phenomenon on interactive digital ads. Although multi-screen apps have not had a large influence on audio and tv ads, there are efforts [15], [16], [17] that show potential impacts of these apps such as: i) an improvement in the interactive methods of content selection by using native accessibility features of secondary devices as smart phones and tablets, and ii) the possibility of delivering synchronized supplementary content to the secondary device screen to enable purchases in response to adverts on the first screen. These approaches focus research on innovative techniques for content recognition and synchronization, APIs creation on top of standard open technology stacks (e.g DLNA and UPnP) to support integration of existing devices, or frameworks creation to enable multiple screen applications by using only one version for all platforms.
- *Content Personalization.* With the huge amount of advertising campaigns and channels available, it is common that a user to receive advertising that is not of his interest. This excess of advertisements can bring the user to a sort of disaffection towards the advertised product, the manufacturer, or more generally, the advertising channel [11]. The key elements of the approaches [18], [19], [20] that seek to maximize the relevance of the delivered advertising are: i) the use of contextual information that allows to place advertisements on the basis of the interests and the activities of the user, ii) the exploitation of data collected from the user behavior to target the advertisements and, (iii) the use of recommend systems to deliver personalized advertising messages based on consumers' behavioral patterns and profiles.
- *Content authentication.* A lot of scientific work has been done in the area of multimedia objects authentication. Authentication in general means deciding whether an object is authentic or not. That is, if it matches a given original object [21]. Different techniques [22], [23] have been proposed in the literature to establish the "equality" of different digital media types (e.g. audio, image or video). However, today a growing scientific and industrial interest is to improve the precision and performance of these techniques.
- *Scalability.* It is a challenge to design a flexible and scalable platform for supporting multi-screen interactive ad delivery. However, recent exponential leaps in computing power, cheap data storage, and open-source software packages for multimedia content analysis has reduced the barrier for exploring novel approaches at scale. Related platforms as presented in [24], [25], [26], try to support the scalability of their systems by addressing factors such as storage, retrieval accuracy, and processing of massive data volumes. An essential factor of a good part of these platforms is to incorporate a SOA-cloud approach. This symbiosis enables optimum orchestration of services combined with optimum elasticity of resources, such as processing power, storage, and number of instances [27]. Moreover, it ensures that all the applications inherit a scalable architecture and can be easily managed and upgraded.

This grouping form the basis to derive requirements for the flexible infrastructure that can support interactive advertising. The requirements introduced below are not exhaustive, but provide a basis for fostering further progress in building up these platforms. The major challenges to be addressed are the following:

- *To support multi-screen ad delivery.* Digital ads across multiple devices allow reaching relevant target groups and boost advertising effectiveness. The platform should be

able i) to understand perceptions and usage patterns of multi-screen users in order to optimize spend allocation and ii) to overcome possible performance problems due to limitations such as the lack of well-structured interfaces between devices, inconvenient connection, and slow network communication speeds.

- *To support multi-tasking services.* Some digital platforms for TV, provide a simple multi-tasking service that display information related to another channel in a small window while watching a program. However, as described in [28], one screen multi-tasking can disturb customer's behavior. We visualize a system able to perform multi-tasking on different devices at the same time, e.g. allowing a user watching a commercial on TV screen to search for additional information about the advertised product via his tablet PC, and finding nearby shops where he can buy it using his smart phone.
- *To deal with digital ads flexibility.* Like other forms of digital advertising, digital ads offer multiple marketing options; for example, local and national sponsorships as well as multiple forms of ad creative suited to many different advertiser types and campaign goals [29]. The platform should be able to identify and classify different ad types, recognize their coverage, and implement appropriate marketing actions for each case.
- *To deal with digital channels fragmentation.* Digital ads are currently distributed in multiple channels such as DTV at all its forms (e.g., iPTV/hybrid/ cableTV, Web TV, etc.), digital radio including streamed satellite radio, pay–per–view service, and others. The platform should support cross channels in an easy and friendly way without requiring any additional software.
- *To implement targeted advertising.* Current digital ad channel diversity may cause that the online delivered programming of a terrestrial station could reach very distant consumers from its home market. In these cases, the platform must be able to show localized ads aimed to target consumers using geographic, demographic or contextual information. Also, the platform has to expand their targeting capabilities to address the digital media based on specific interests and behavioral patterns. In [30], the authors propose an additional option where the consumers can be targeted based on content, keywords and hashtags shared on social networks.
- *To cope with complex usage scenarios.* A commercial may be announced at a time the user cannot take appropriate actions (for example, while driving a car). For these cases, we envisioned a platform that incorporates speech recognition technology to perform actions with the voice. This technology can be used also for ingenious custom interactive campaigns, as well as branded digital audio online content.
- *To recognize digital ads independent of its duration.* Currently there is no single, widely accepted "standard" length for ads spot breaks. For digital audio, the Interactive Advertising Bureau (IAB) reports that there is a wide range of spot break lengths, where the majority of audio ads delivered in–stream are 60, 30 or 15 seconds long. Regardless of the spot break length and the ads duration, the platform must be able to recognize all ads and execute the corresponding marketing actions.
- *Easy to extend to new digital ads.* The platform should have the ability to include new digital ads with a minimum effort required for extensions implementation. This means that no major platform alterations should be needed when the management process is applied to a new digital ad. This feature should become a key design criterion for any platform.
- *To support services, not structures.* The platform must be coped with evolving technology, offering the available capabilities through flexible components and, where appropriate, through APIs.

All requirements above described are the base of our generic architecture described in the next section.

V. General Architecture

The philosophy used in the design of the general architecture was to create a platform that supports a great part of the system requirements described in the previous section. The platform deploys *Software as Services* (*SaaS*) interfaces & APIs, to solve the difficulties faced by the three primary groups of people involved in interactive advertising: ad consumers, advertisers and marketers, and developers and marketing designers. High–level modules of the architecture are illustrated in Figure 1, where principal modules are represented by a solid rectangle and sub-modules are represented by a dotted rectangle. The solid line represents the main process flow; whereas data access is shown as a dotted line on the figure.

We envisaged that different multi-screen interactive ad delivery systems may group two or more components into a single component or may not implement certain components at all. Architecture features will be explained in this section.

A. Main and Second Screen

Communication between first and second screen can be performed directly or by information extraction from a network source. A direct communication protocol is DIAL[9], which uses a REST service to invoke commands. It protocols such as UPnP[10] or DLNA[11] could be used if the application does not require the support of *screen mirroring*. Other options to communicate first and second screens in a local network are: WebSockets[12] or HTTP Streaming [31]. The employed way of communication depends on the variations of the available equipment. An overview of equipment requirements for each case can be consulted in [3].

On the second screen the following components are neccesary: i) the *Device Connection* component, which ensures the second screen content can be accessed from the first

[9]http://www.dial-multiscreen.org/home
[10]http://www.upnp.org/
[11]http://www.dlna.org/
[12]https://tools.ietf.org/html/rfc6455

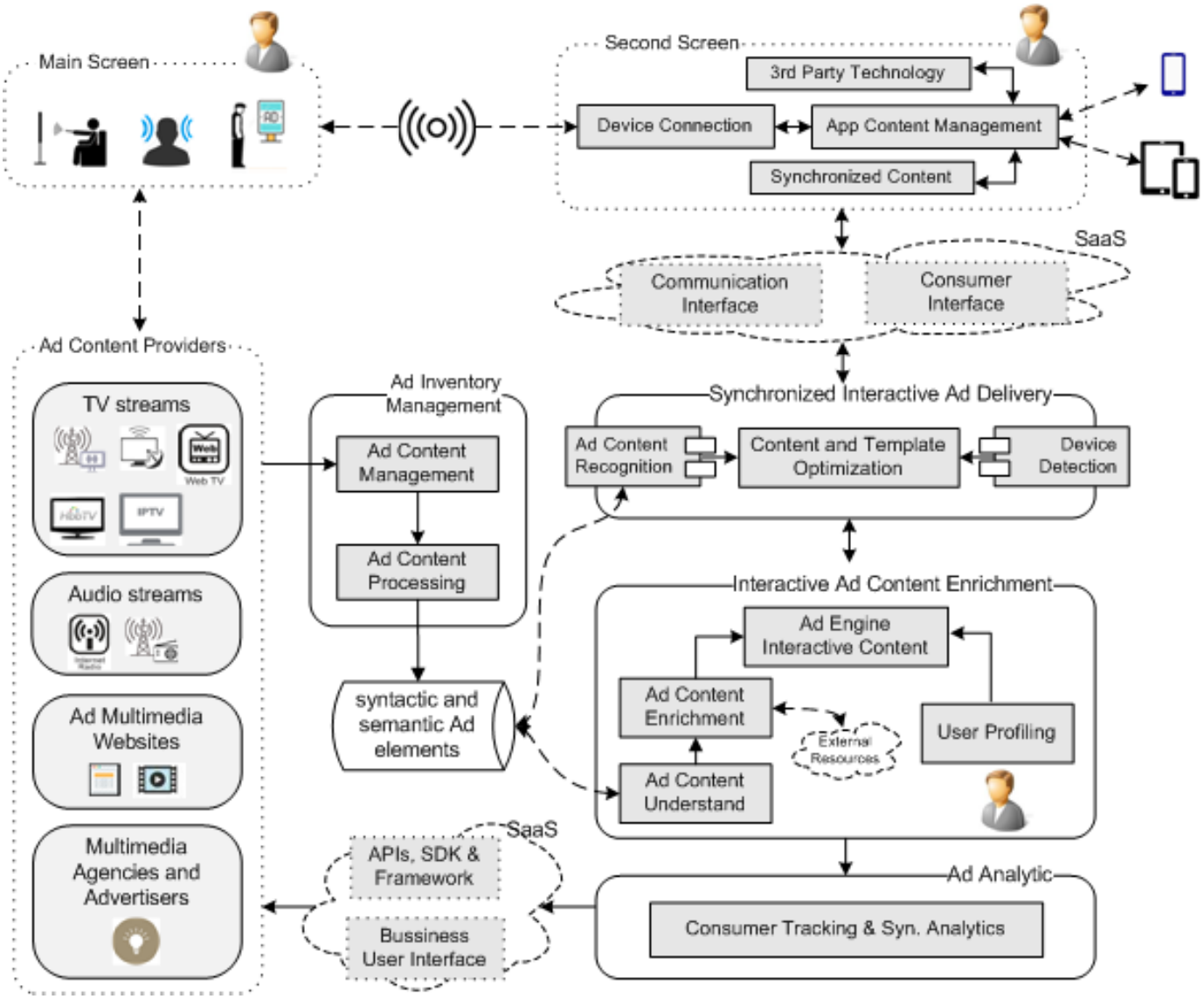

Fig. 1. General Architecture of the Interactive Ad Delivery Platform.

screen, ii) the *3rd Part Technology* component, which implements additional technologies for supporting complex usage scenarios, e.g. voice commands, iii) the *App Content Management* component for interactive ad content display, and iv) the *Synchronized Content* component used to identify ad content and to fetch related information from other sources.

Ad content identification in the platform can be achieved using the following techniques: Linking Devices, IP Sync, Automatic Content Recognition (ACR), etc. Technical works [32], [33] show how ACR technologies (watermarking or fingerprinting) can be used to provide automatic recognition and synchronization. These techniques could be coupled to our aims, since the platform will allow broadcast payload manipulation and fingerprint recognition infrastructure.

B. Ad Inventory Management

This module is responsible for managing extracted advertising automatically from multiple *Ad Content Providers*. The offline tasks to be performed by this module are: i) to monitor and storing digital ads that need to be identified, enhanced, and delivered to consumers on multiple companion devices, and ii) to analyze digital ads in order to extract both syntactic and semantic elements.

The *Ad Content Management* component, can use different techniques to automatically collect digital ads. A suitable option is to use Web scraping techniques to develop an autonomous system that gathers video and audio information from ad Web pages [4]. Other approaches attempt to monitor and recognize a particular ad from broadcaster streams or video databases [34], [35].

Once an ad is collected, the next step is storing it. The

analysis of multiple ad management tools allowed us to observe that progress on this field has led to a number of concrete well supported implementations. These tools allow to structure the storage at different levels: *campaign*, *ad set* and *ad*. A campaign can contain one or more ad sets, whereas an ad set has one or more ads. Each ad contains its design or ad creative (e.g., video, audio, text or images). We envisage a fourth level where each ad visual element can be split and stored as well.

The *Ad Content Processing component* is responsible for extracting features from audio or video ads in order to efficiently filter, rank, index and retrieve ads. Depending on the level of content processing required in the application, this component can use algorithms at two levels: *ad file* and *ad content*.

The two leading ACR technologies, watermarking and fingerprinting can be used to analyze the features of an *ad file*. Watermarking consists in inserting a unique pattern in the content stream to allow its identification. Fingerprint, instead performs the comparison based on hash values obtained from the analysis of unchanged characteristics in the audio and video frames. One of the key characteristics of fingerprint technology is that it does not modify the original content. Also, digital fingerprints are highly compressed when compared to the original source file and can therefore be easily stored in databases for later comparison [33].

We agree with [4] that the *ad content* analysis involves the need of extracting syntactic and semantic elements from digital ads. Syntactic elements refer to those ad elements existing in digital ads such as images, visual logos and text, whereas semantic elements are usually high–level concepts, which related to viewer ad content understanding, i.e., visual concepts or textual keywords. Techniques such as the Histogram of Oriented Gradients image descriptor and a Linear Support Vector Machine could be used to train highly accurate object classifiers [36]. Deep Neural Networks can also be used to identify objects in an accurate manner [37]. These objects detection and recognition techniques, could extract syntactic elements in an automatic way. As analyzed in [4], the detection of object concepts and context concepts can be resolved using a per-sample multi kernel approach [38]. This technique offers a main advantage over other approaches because it works on sparse samples.

C. Interactive Ad Content Enrichment

This module is the key element for enriching digital ads with interactive, contextual and personalized content that can be deployed on multiple companion devices. The *Ad Engine Interactive Content* is the dynamic component that enables real–time decisions and directs relevant advertising and content to customers by mapping ad content with user data.

A recommendation system can solve this task, automatically discovering items best matching users' preferences, interests and needs. However, as analyzed in [39], traditional recommender systems do not fully meet the needs of e–commerce settings in which users are not focused on the kind of items offered to them. Its authors propose that the functionalities of a typical recommendation system can be extended by using semantic modeling. While the recommendation algorithm proposed in [39] takes advantage of an internal reasoning about ads and user profiles, and, of an effective hybrid combination of content-based and collaborative filtering to automatically compose interactive applications, there is no inclusion of user context to offer personalized commercial functionalities.

In our context, individual and group users can interact with several companion devices at the same time. Therefore, it is crucial that recommendation algorithms support cross-context personalization. A possible solution is to provide a domain independent modeling of users, devices and contexts [40].

The *Ad Content Understanding (ACU)* component supports ad classification in one or various predefined classes based on their content. There are many ways to perform video&audio classification. One way is through digital signal processing. Another, and in our opinion, one of the most promising approaches is based on the analysis of perceptual features i.e., syntactic and semantic elements as described in the previous step. Most existing approaches attempt to perform statistical [41] or semantic-based techniques [42] to associate a set of perceptual features to ad production types and advertised products/services. An advantage of semantic–based technique is that the connection between the annotation subject (in this case, the ad content element) and the annotation object (recognized concept to be displayed on the corresponding element) can be expressed in a machine–processable language. This feature may enable automatic content discovery.

For our purposes, the *ACU* component has to support and to extend current ad classification approaches, taking into account other aspects such as: ad coverage, ad formats, trade agreements, etc. To achieve this goal the component may use an ontologies network.

The interactive content ad enrichment process is supported by the *Ad Content Enrichment* component, which can extract additional information from external resources or use metadata supplied by the advertising agency. In a previous work [43], we show the feasibility of using external resources to enrich the content of electronic programming guides. An improvement to this proposal would be the use of the W3C Media Fragment URI specification to enable direct access to ad elements into video and audio without having to retrieve the entire resource.

To be effective, the platform needs to ensure that the ad content and interactive messages are relevant for the consumer who sees it. The *User Profiling* component aims to keep and manage all user's profiles including information related to interests, disabilities, preferences and devices used on the platform. For example, users interaction to the platform may reveal their interests, whereas the information extracted from companion devices may provide information on the location and context. Several approaches have been adopted for this purpose, ranging from indexed features vectors, ratings matrices and demographic features sets, decision trees, neural networks based classifiers and semantic networks.

D. Synchronized Ad Delivery

The key step to a successful multi–screen marketing campaign stands on creating a unique experience for each device. To achieve this goal, the *Synchronized Ad Delivery* module relies on the *Content and Template Optimization (CTO)* component. From a technical view point, this component tries to provide the best second–screen user experience (UX) possible. Two principles have been considered to offer a personalized UX within the platform: i) to consider a personalized experience to be agile, such that it can be easily steered in directions that users choose, and ii) to simplify the user experience using "native" template interfaces for each device, avoiding dealing with the pain of squeezing all of them into one unique template.

The *Ad Content Recognition* component supports the first goal of the *CTO* component, which is detecting the actions performed on the user companion device. This component listens to the actions performed on the device and extracts a feature vector of descriptive metadata, i.e. metadata that describes the unique characteristics of an audio or video ad file. Extracted metadata can be divided into syntactic and semantic descriptors. Upon ad identification, the syntactic descriptors detected by the *Ad Content Recognition* component include information such as the channel ID, program ID and time stamp info retrieved from ad database and/or ad elements such as images, visual logos or text. Semantic descriptors can extract two types of ad related visual concepts: product (object concept) and its context information (context concept), such as a person using the product or general scenes depicting the use of the product.

The *Device Detection* component is responsible for detecting the type of companion device used by the user. The aim of this task is to avoid the need of performing *responsive design* optimization. As described in [44], *adaptation design on server* offers benefits such as: i) optimization of not only the HTML code, but also the assets related to it, ii) creation of substantially different layouts on different screens, and iii) optimization of front–end performance by relying on the back–end high performance.

E. Ad Analytic

The aim of this module is to create an analytic strategy that helps marketers to understand their customers in a holistic way. For this purpose, the *Consumer Tracking & Syn. Analytics* component must implement algorithms to track a specific consumer across their devices and to implement a universal metric that be consistent across multiple devices.

A report of MIT [45] shows that technology companies have begun to develop technology that uses website logins, shared Wi-Fi connections and search histories to connect a single user to their multiple devices. Other methods include collecting anonymous desktop Web, mobile Web and mobile app user interactions to make the association. These approaches can be used as a base to build own technology to help advertisers reach customers across multiple second devices.

Regarding cross-screen analytic metrics, we see yet a strong need for measurement platforms that will allow marketers to understand intent harvesting, generation and creation in an integrated manner.

VI. Prototype Applications

The proposed platform is being applied and tested in different applications, two of which are briefly described in this section.

Identification of Vehicle Advertisements

The aim of the two first applications is to test different audio and video recognition algorithms to identify ads through a second screen device. Figure 2 shows the technology used to implement the different components of the architecture.

The scenario begins with two *Ad content providers*: audio and videos about different vehicle brands extracted from YouTube. The *Ad Content Management component* uses a Python server to store ad metadata automatically extracted from AdsoftheWorld Web page[13]. This Web page ranks the advertisements by topic, where each ad page contains information such as: ad name, video duration, advertising agency, or the link to YouTube file. All information is extracted using Web scraping techniques.

In order to extract the video and audio ad features, the *Ad Content Processing component* implements two algorithms respectively:

- *Perceptual hash functions* [22], which are based on extracting robust features from multimedia data. These features are closely related with the content, regardless of the binary representation. Therefore they can resist a certain range of legitimate distortion [21]. This type of algorithms allows comparing two images using a similarity measure. The comparison is done between short binary strings that represent extracted features from the images.
  In [46] we perform a benchmarking process of the four most know, available, and free perceptual image hashing algorithms to determine their efficiency. The evaluation results suggested that employing just a sample frame to identify a video is possible, due that the obtained success rate was upper than 98.22% using Perception Hash (PHASH) with Hamming Distance of 6.
- *Algorithms based on audio features extraction* [23], are presented as an alternative to perceptual hashing algorithms. These algorithms allow increasing the scope of recognition and extending the number of multimedia sources such as: Internet radio, Internet audio streaming and broadcasting radio signals. Currently, audio fingerprint technologies have attracted attention since they allow the identification of audio independently of its format and without the need of meta-data or watermark embedding [47]. Also, they have a good accuracy when

[13] http://adsoftheworld.com/

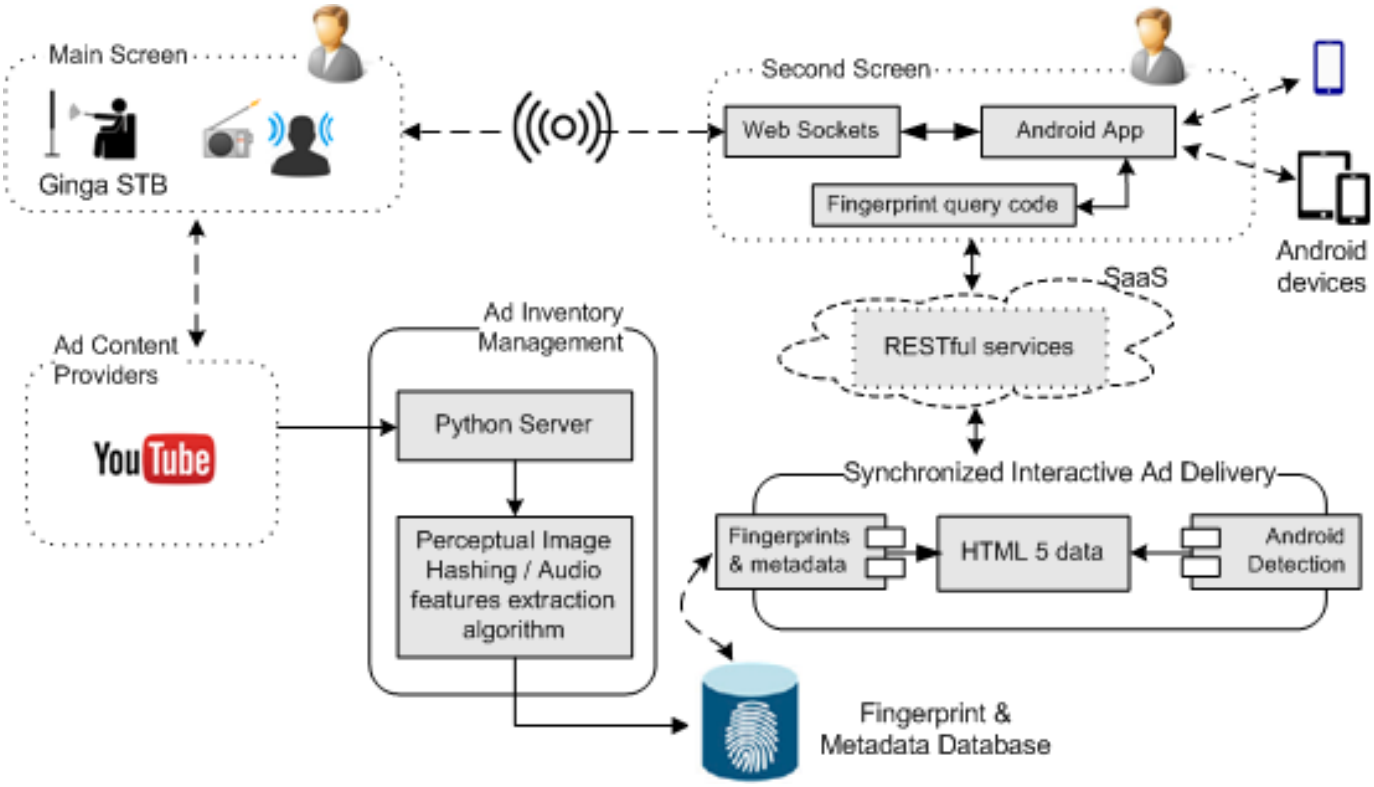

Fig. 2. Ad Recognition Prototype.

the audio has degradation, noise addition and format transformation.

The fingerprint system proposed in [48] shows the feasibility of fingerprint for advertising. This work remarks many important characteristics in the parameterization of an advertising fingerprint algorithm. Also, It considers features related to voice, music and noise in order to obtain high accuracy around 99% for short audio samples in broadcast monitoring.

The result from this process is a database of metadata and fingerprint references.

To recognize an advertisement, the user takes a sample of audio or video through an android app implemented in the *App Content Management component*. In order to enable the connection between the android device and the Ginga Set–top box (STB), a Web server has been deployed to send a screen capture from the STB when the user requests it through the android device. These actions are monitored by the *Ad Content Recognition* component. The android device connects via sockets to the STB and via REST services to Python server that contains the ad metadata and fingerprint references. The *Synchronized Content component* is the responsible for sending the fingerprint query to retrieve the advertising and its metadata. The results of the algorithms implemented for identifying video and audio ads are summarized as follows:

- *Perceptual hash functions*: The evaluation process consisted in count the success rate in a true table. The database employed during the benchmarking process was populated with 246145 video frames, obtained from 165 videos of automobile ads; 30000 queries were done randomly picking 1 frame each time from a dataset composed of the same videos in the database, and extra videos that were not trained. All the videos were previously subject to changes of resolution, codec, quality and format. The total number of available frames was 4644307. The proposed Ads Video Identification from one or several frames is highly robust, showing a success rate higher than 98.22%. Obviously using more than one frame will increase the success rate. Also, it worth comment that the identification process gives results in real time (time processing average of 0.1034 seconds per frame, employing 1 processor of 2.7Ghz). Nevertheless, the errors found were due to the ambiguity of certain frames, which are the same in different videos. E.g. a frame with just a black background.
- *Algorithms based on audio features extraction*. The fingerprint system [48] takes a set of 600 advertisements that are normalized to raw format with PCM 16-bit audio codec. The fingerprint feature extraction is based on highest power peaks, and takes two points: an anchor peak and another in an area called target zone. This method makes the algorithm robust to degradation because peaks are enough robust in order to survive to noise and postprocessing. During peaks extraction audios are framed to 8 milliseconds and each frame had an overlapping factor of 50%. Finally, the system database is extended around of

240 thousand of fingerprints. The evaluation process was made taking 363 random audio samples and 3 seconds in duration; as a result the algorithm has a success rate of 99.7%, a false negative rate of 0.3%, and a false positive rate of 0%.

Once the application recognizes an advertising, the *Content and Template Optimization component* is responsible for generating the HTML5 code with the ad metadata, which are sent to the android device.

The demos and the results of the benchmarking process used to test the ad recognition apps for audio and video are available at http://190.15.132.90:9051/fingerprint and http://190.15.132.91:9001/PIH, respectively.

VII. Conclusions and Future Work

In this paper, we outlined desirable requirements and described an interactive ad platform based on key factors derived of related works and own experiences. The proposed platform intends to make use of a set of technologies to complete a given task. To the best of our knowledge, this will be the first time that these technologies are put together to capture people's attention on specific services, topics or products by using interactive ad visual and audible experiences.

To illustrate the flexibility of our platform, we have described two basic applications that allow identifying an ad using audio or video recognition algorithms through a second screen device. As future work, we would like to explore the potential of our platform in other applications that require ad content enrichment or ad analytic results.

Acknowledgments

The work presented in this article is part of a research project called "Use of semantic technologies to analyze multimedia content broadcasted by digital television" supported by the Department of Research of the University of Cuenca (DIUC).

References

[1] Google, "The new multi-screen world: Understanding cross-platform consumer behavior," www.services.google.com/fh/files/misc/multiscreenworld_final.pdf, accessed: 2016-05-01.

[2] Microsoft, "Cross-screen engagement," http://advertising.microsoft.com/international/cross-screen-research-report, accessed: 2016-05-01.

[3] J. Klein, J. Freeman, D. Harding, and A. Teffahi, "Assessing the impact of second screen," Technologia in association with DTG and i2 media research, Tech. Rep., 2014.

[4] J. Wang, M. Xu, H. Lu, and I. Burnett, "Activead:a novel framework of linking ad videos to online products," *NeuroComputing (InPress)*, 2016.

[5] Y. Blanco-Fernández, M. López-Nores, J. J. Pazos-Arias, and M. I. Martín-Vicente, "Exploring new ways for personalized e-commerce through digital TV," in *Semantic Hyper/Multimedia Adaptation - Schemes and Applications*. Springer, 2013, pp. 181–201.

[6] J. D. Leckenby and H. Li, "From the editors: Why we need the journal of interactive advertising," *Journal of Interactive Advertising*, vol. 1, 2000.

[7] D. Zahay, *Digital Marketing Management: A Handbook for the Current (or Future) CEO*, ser. Digital and social media marketing and advertising collection. Business Expert Press, 2015.

[8] E. Thomas, J. Z. Pan, S. Taylor, Y. Ren, N. Jekjantuk, and Y. Zhao, *Future Internet - FIS 2009: Second Future Internet Symposium, FIS 2009, Berlin, Germany, September 1-3, 2009*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ch. Semantic Advertising for Web 3.0, pp. 96–105.

[9] S. Albayrak, J. Glass, B. Karatas, S. Marx, T. Schmidt, and F. Sivrikaya, *Towards Semantic IPTV*. River Publishers, Jan. 2011, pp. 197–130.

[10] H. J. Müller, F. Alt, and D. Michelis, "Pervasive advertising," in *Pervasive Advertising*. Springer London, 2011, pp. 1–29.

[11] L. Carrara and G. Orsi, "A new perspective in pervasive advertising," DCS, Tech. Rep. RR-11-07, 2011.

[12] S. Bellman, A. Schweda, and D. Varan, "A comparison of three interactive television ad formats," *Journal of Interactive Advertising*, vol. 10, 2009.

[13] B. Katz, *Interactive TV in the UK: New advertising opportunities in comparison*. Diplom.de, 2001.

[14] A. Khanna, "Integrated marketing communication creative strategy from idea to implementation by robyn blakeman," *Global Journal of Enterprise Information System*, vol. 4, no. 1, 2012.

[15] H. Simon, E. Comunello, and A. V. Wangenheim, "Enrichment of interactive digital tv using second screen," *International Journal of Computer Applications*, vol. 64, no. 22, pp. 58–64, February 2013.

[16] T.-H. You, Y.-J. Wu, C.-L. Lin, and Y. Chuang, *Enhancing People's Television Experience by Capturing, Memoing, Sharing, and Mixing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 510–518.

[17] C. Krauss, L. Bassbouss, S. Pham, S. Kaiser, A. Stephan, and S. Steglich, "Challenges for enabling targeted multi-screen advertisement for interactive tv services," in *Fourth W3C Web and TV Workshop*, 2014, pp. 13–15.

[18] L. Ceccaroni and X. Verdaguer, "Tv finder: una aproximación semántica a la televisión interactiva," in *Proceedings of the workshop on Ubiquitous computation and ambient intelligence of the Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2003), Donostia, Spain*, 2003.

[19] R. Sotelo, M. Juayek, and A. Scuoteguazza, "A comparison of audiovisual content recommender systems performance: Collaborative vs. semantic approaches," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*, June 2013, pp. 1–5.

[20] V. Saquicela, M. Espinoza-Mejía, J. Mejía, and B. Villazón-Terrazas, "Reduciendo la sobrecarga de información en usuarios de televisión digital," *in Proceedings of the Workshop on Semantic Web and Linked Data*, 2014.

[21] C. Zauner, "Implementation and benchmarking of perceptual image hash functions," Ph.D. dissertation, University of Applied Sciences Hagenberg, 2010.

[22] L. Weng, "Perceptual multimedia hashing," Ph.D. dissertation, Katholieke Universiteit Leuven, 2012.

[23] R. M. C. Segundo and C. A. S. Santos, "Second screen event flow synchronization," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB 2013, London, United Kingdom, June 5-7*, 2013, pp. 1–7.

[24] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1725–1732. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2014.223

[25] J. Carretero, J. G. Blas, R. K. L. Ko, P. Mueller, and K. Nakano, Eds., *Algorithms and Architectures for Parallel Processing - 16th International Conference, ICA3PP 2016, Granada, Spain, December 14-16, 2016, Proceedings*, ser. Lecture Notes in Computer Science, vol. 10048. Springer, 2016. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-49583-5

[26] A. Araujo and B. Girod, "Large-scale video retrieval using image queries," *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[27] P. Wik, "Thunder clouds: Managing soa-cloud risk - part i and ii," *Service Technology Magazine*, 2011.

[28] J. Jeong, H. Hong, and D. Lee, "Ontology-based automatic video annotation technique in smart TV environment," *IEEE Trans. Consumer Electronics*, vol. 57, no. 4, pp. 1830–1836, 2011.

[29] Interactive Advertising Bureau, "Platform status report: A digital audio advertising overview," IAB, Tech. Rep., 2010.

[30] G. P. Parameswaran and R. Vanjani, "Multi-screen advertising unveiling cross platform advertising solutions," Wipro Council for Industry Report, Tech. Rep., 2013.
[31] S. Loreto, P. Saint-Andre, S. Salsano, and G. Wilkins, "Known issues and best practices for the use of long polling and streaming in bidirectional HTTP," RFC 6202, Internet Engineering Task Force, 2011. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6202.txt
[32] D. Milano, "Content control: Digital watermarking and fingerprinting," Rhozet, Tech. Rep., 2012.
[33] Civolution, "Automated content recognition creating content aware ecosystems," Civolution, Tech. Rep., 2012.
[34] P. Duygulu, M. Chen, and A. G. Hauptmann, "Comparison and combination of two novel commercial detection methods," in *Proceedings of International Conference on Multimedia and Expo, ICME*, 2004, pp. 1267–1270.
[35] A. Shivadas and J. M. Gauch, "Real-time commercial recognition using color moments and hashing." in *Computer and Robot Vision, 2007. CRV '07*. IEEE, 2007, pp. 465–472.
[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2005, pp. 886–893.
[37] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE, 2014, pp. 2155–2162.
[38] J. Yang, Y. Li, Y. Tian, L.-Y. Duan, and W. Gao, "Per-sample multiple kernel approach for visual concept learning," *J. Image Video Process.*, pp. 2:1–2:13, 2010.
[39] M. López-Nores, Y. Blanco-Fernández, J. J. Pazos-Arias, A. Fernández-Vilas, and M. Ramos-Cabrer, "Automatic provision of personalized e-commerce services in digital TV scenarios with impermanent connectivity," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12 691–12 698, 2011.
[40] L. Aroyo, P. Bellekens, M. Björkman, G.-J. Houben, P. Akkermans, and A. Kaptein, "Sensee framework for personalized access to tv content." in *EuroITV*, vol. 4471. Springer, 2007, pp. 156–165.
[41] J. Wang, L. Duan, L. Xu, H. Lu, and J. S. Jin, "TV ad video categorization with probabilistic latent concept learning," in *Proceedings of the 9th ACM SIGMM International Workshop on Multimedia Information Retrieval, Augsburg, Germany, September*, 2007, pp. 217–226.
[42] J.-H. Lee, J. Ha, J.-Y. Jung, and S. Lee, "Semantic contextual advertising based on the open directory project," *ACM Trans. Web*, vol. 7, no. 4, pp. 24:1–24:22, 2013.
[43] V. Saquicela, M. Espinoza-Mejía, K. Palacio, and H. Alban, "Enriching electronic program guides using semantic technologies and external resources," in *XL Latin American Computing Conference, CLEI, Montevideo, Uruguay*, 2014, pp. 1–8.
[44] eZ Blog, "The paths to the best multi-screen user experience," http://ez.no/Blog/The-Paths-to-the-Best-Multi-screen-User-Experience, accessed: 2016-05-01.
[45] MIT, "New technology for tracking consumers across devices grows results," http://www.digilant.com/whitepaper/techreview/, accessed: 2016-05-01.
[46] F. Vega, J. Medina, D. Mendoza, V. Saquicela, and M. Espinoza, "A robust video identification framework using perceptual image hashing," in *43th Latin American Computing Conference, Lat. Am. Symp. on Computer Graphics, Virtual Reality, and Image Processing*, 2017.
[47] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A review of audio fingerprinting," *The Journal of VLSI Signal Processing*, vol. 41, pp. 271 – 284, 2005.
[48] J. Medina, F. Vega, D. Mendoza, V. Saquicela, and M. Espinoza, "Audio fingerprint parametrization for multimedia advertising identification," 2017, submitted to IEEE Second Ecuador Technical Chapters Meeting.