2012 IEEE International Conference on Multimedia and Expo Workshops

Web-based Augmented Reality Video Streaming for Marketing

Ville Valjus, Sari Järvinen, Johannes Peltola
ICT, Media Technologies
VTT Technical Research Centre of Finland
Oulu, Finland
ville.valjus@vtt.fi, sari.jarvinen@vtt.fi, johannes.peltola@vtt.fi

***Abstract*—This paper presents an Adobe Flash-based augmented reality video streaming application and its practical use in web marketing. The application enables augmenting the content of a webcam view by adding video content to it. Aller Media, a Nordic media company, used the application for advertising two Finnish movies with promotional video content. We have examined combining of conventional print media and digital media and the suitability of the augmented reality video streaming application for the web environment and marketing purposes. Additionally, we measured the technical performance of the application. The feedback from Aller Media and the end-users indicates that the application is useful for marketing purposes. In addition, the results show that the application is well suited for web environment as its performance is sufficient and as the distribution is more efficient compared to desktop or mobile applications.**

Keywords-augmented reality; Adobe Flash; video streaming; marketing

I. Introduction

The web is an effective channel for marketing and advertising. In many cases it competes with conventional print media and during recent years it has gained ground from print media. Instead of this rivalry, print media and web-based digital media should be combined in order to complement each other. The *augmented reality* (AR) technology is a potential solution for this situation.

AR brings added value to web marketing through interactive applications. Augmented reality is a form of *virtual reality* (VR) [1, 2]. The Azuma definition [1] specifies AR as systems having the following characteristics:

1) Combination of real and virtual: The AR environment is based on the real world around the user. The user can see the real world, but with synthetic objects embedded in it. AR thus seeks to enrich reality rather than to completely replace it [3].

2) Interactive in real-time: The user is able to influence events taking place in the AR environment.

3) Registered in 3D: The synthetic objects embedded in the environment must be true 3D objects instead of simple 2D overlays.

AR applications have traditionally been developed for desktop or mobile environments, such as the application for the children's storybook of "The Dibidogs and the Lost Princess" [4] presented in Fig. 1. The applications require installation on the target device prior to use, which can be a disincentive for many potential users who may hesitate to install additional software on their devices. Web-based applications, however, offer a solution to this problem as the AR content can be presented as part of a web page without any additional installations, thus enabling much more efficient distribution of applications. In marketing this means delivering the promotional content faster and attracting more users leading to more potential customers for the products.

This study presents a web-based AR application for marketing purposes. The application utilizes Adobe Flash platform and its core idea is to enable video content to be presented on top of a marker visible in a webcam feed. The application was developed in a commercial project for Aller Media, a leading entertainment and TV magazine publisher in Finland, and it was used for promoting two Finnish films, "Veijarit" and "Vares – Pahan suudelma", with markers printed on the magazine Seiska. The purpose of the application is to combine conventional print media and digital media by adding interactivity to the magazine and thus enrich the reading experience.

In this paper we present few current AR-based marketing applications and reflect our own solution against them. We also discuss the requirements and the implementation of our video streaming application, and the augmented reality library ALVAR utilized in the study. Furthermore, we describe the suitability of the application for marketing and the web environment and what kind of performance it is capable of achieving.

Figure 1. The Dibidogs and the Lost Princess AR application.

978-0-7695-4729-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMEW.2012.63

The structure of this paper is the following. Augmented reality applications are presented in Section II. Section III discusses the developed web-based video streaming application while the testing of the application is presented in Section IV. Discussion and conclusion are presented in sections V and VI, respectively.

II. AUGMENTED REALITY MARKETING SOLUTIONS

This chapter discusses augmented reality applications presenting a basic structure of an AR application and few examples of marketing AR applications implemented with different technologies.

The basic structure of an AR application (Fig. 2) can be described as follows. Firstly, a camera is used to capture the user's environment for further processing. The capture from the environment is searched for a predefined element, a marker, which can be, for example, a black-and-white fiduciary marker (Fig. 3), a custom image or a 3D point cloud representing some object in the real environment, such as a building. In all cases the application needs to be taught to recognize the markers so that the application knows what to identify. Once the application has recognized the markers from the camera capture, it uses various algorithms to identify the markers and to determine their orientation and position. The application then superimposes synthetic content on top of the camera capture. The content is usually located and orientated in relation to a visible marker by using the marker information. The content can thus be positioned with a very high degree of accuracy, thus enhancing the illusion that the synthetic content is actually part of the environment.

Augmented reality has become a popular tool for marketing applications. For example, virtual dressing rooms enable the shopper to try on different clothes virtually without physically stepping foot in the clothes store. The user can get an impression how the products would look when worn in the real world. Virtual dressing rooms are not just a marketing tool. They also bring additional value to customers' shopping experience. For instance, establishments such as the online fashion boutique Tobi (http://www.youtube.com/watch?v=rn_iPjGKd0M) and the department store chain J. C. Penney Company, Inc. (http://www.youtube.com/watch?v=fhjuZMEJ4-U) provide virtual dressing rooms on their web pages. Both of these dressing rooms enable customers to try on various items from the clothing range by overlaying images of the selected clothes on the webcam view.

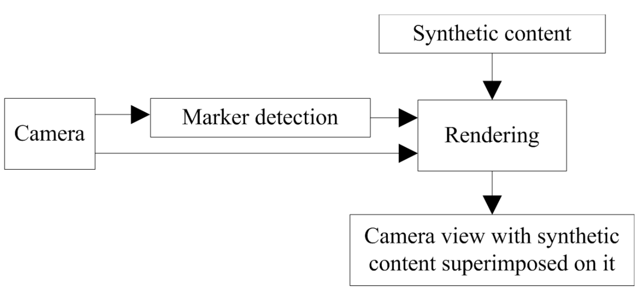


Figure 2. The basic structure of an AR application.

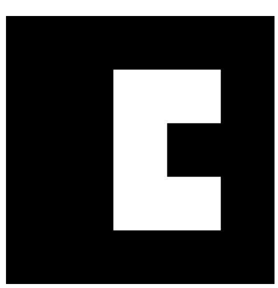


Figure 3. A black-and-white fiduciary marker.

The versatility of the toy industry provides a vast range of possibilities for AR applications. LEGO Group (http://www.youtube.com/watch?v=mUuVvY4c4-A), for instance, has an interactive terminal for revealing the contents of LEGO packages. The application uses a camera to detect the LEGO box and a display screen to show a fully-assembled 3D model on top of the box being shown to the camera. This allows the customer to examine the assembled product before making their purchase decision.

Toyota markets one of its models, iQ, using an AR application (http://www.youtube.com/watch?v=1_7NW_u3VFo). The application enables the user to get a feel of the car's agility and view its interior. When a marker is shown to the webcam, the car is animated to be driven on a road, and the user can steer the car by tilting the marker. Alternatively, the user can examine the car's interior via an exploded view.

The Harry Potter film series has also been marketed using augmented reality. The Harry Potter theme park, The Wizarding World of Harry Potter, has been advertised by developing an AR application (http://www.harrypotter3d.com) which uses a theme park related map as the marker and superimposes different locations and characters on the map, thus allowing the user to familiarize themselves virtually with the theme park. The Harry Potter and the Deathly Hallow - Part 2 videogame can also be experienced via AR (http://harrypotter.ea.com/?page_id=464) as the videogame box acts as a marker. The application is relatively simple, featuring an animation of Harry Potter casting spells in 3D.

Electronics differ from the previous product categories in that technical specifications and functionality are often the key aspects. This can be seen in the Olympus' AR advertisement (http://www.getolympus.com/pen/pen3d.asp), in which the PEN E-PL1 system camera is promoted. The web page allows the user to interact with the camera and to simulate product features, such as taking pictures, shooting video, trying the flash and removing the lens. All this is available via marker detection technology and a 3D model of the camera superimposed on the marker held by the user.

These applications are just a few individual examples from a large collection of AR marketing applications developed. Currently many of the AR marketing applications are web-based but very few, if any, of them are focused on video streaming. Thus, our web-based video streaming application brings new innovation to the field of augmented reality.

III. WEB-BASED VIDEO STREAMING APPLICATION

A. Background of Our Work

The initiative for our work came from the media company Aller Media and their need for web-based video streaming AR application (Fig. 4) for marketing purposes. They printed markers on their magazines and their readers were able to watch film trailers and other related video clips from a website using the magazine and a webcam. Combining conventional print media and digital media it is possible to add interactivity to the magazine enriching the reading experience. The requirement specifications and functional design were carried out in co-operation with the company.

B. ALVAR Flash Library

VTT has previously developed an AR software library called ALVAR (A Library for Virtual and Augmented Reality) [5]. The library can be used for developing AR applications for desktop and mobile environments, and we chose to use it also for our web-based video streaming application. However, the existing version of ALVAR can only be used for developing desktop and mobile AR applications and thus could not be used as-is. ALVAR therefore had to be first converted to a web-based technology. In this study, the chosen technology for this purpose was Adobe Flash as this platform is currently a highly popular technology in the web environment, having a support for webcam capture and being able to display content on multiple operating systems on computers and mobile devices. The converted software was thus named ALVAR Flash Library. The conversion was done using Adobe Alchemy which is an ongoing research project by Adobe allowing developers to compile C and C++ source code to byte code optimized for ActionScript Virtual Machine (AVM2).

The main functionality of ALVAR is to detect black-and-white fiduciary markers from input image data and to return information about the found markers as a response. The process can be described as follows. Firstly, an input image must be given. The input must be a pointer to an unsigned character array. Secondly, ALVAR uses its algorithms to search the input image for any visible markers. Thirdly, information such as pose, rotation, translation and the corner points of detected markers are provided for further processing.

Figure 4. Web-based video streaming application.

The ALVAR Flash Library interface was implemented in one C++ source code file containing the two functions, one for initialization and another for processing individual video frames. ALVAR's functionality was included in the code as static header files. ActionScript related functionality was similarly added by including one Alchemy related C++ header file.

At first, the interface contained only one function: to process video frames by receiving a copy of the image data and related parameters. As passing a multichannel high resolution image through the interface as actual values, and thus copying it into another memory location, caused a significant delay, a shared memory space for storing the image data was adopted. Consequently, there was a need for an initialization function for allocating the memory space. Receiving input parameters was also added to this function, as it is not necessary to receive them separately for each frame.

In addition, grey scale images were originally calculated so that each color component of each pixel was looped through, and a grey scale value for that pixel was calculated from the combination of all color components. After the optimization, only part of the pixels was looped through. As the width and height of the input image is divisible by two, the total resolution from multiplying these two is divisible by four. It was therefore more efficient to stop at every fourth pixel and, specifically, at its last color component, and calculate the grey scale value based on the previous three pixels and their color components.

C. Requirements

The core functionality of the Flash video streaming application was to identify markers from webcam video, to stream a video, and to show the video and additional objects in a 3D plane on top of the marker in such a way that the video responds to the marker's movement. The purpose of the additional objects is to enrich the surroundings of the video and to enhance the 3D environment by positioning objects at different depths. The application had to support multiple markers and videos, but needed to display only one video at a time. In order to render 3D objects on top of a marker, it was necessary to utilize a third party Flash 3D engine and Papervision3D was chosen for this task as it seemed to be the most promising choice for the application.

In cases where there are multiple visible markers, but only one video can be played on top of one marker at a time, it has to be decided which video will be played and on which marker. For this, a configuration file is required (Fig. 5). The file must define the markers that the application has to identify, and which video should be played for each marker. The configuration file must also define a timer delay for stopping playback and changing the currently used marker.

In the first case, the timer starts running when there are no visible markers defined in the configuration file. The configuration file thus prevents breaks in the video caused by possible lapses in marker detection in individual frames. In the second case, the timer affects the decision regarding which marker is currently active by waiting for the timer to run out before assigning a higher priority marker as the active one. The active marker is the marker on top of which the video is to be shown. The marker priority is defined by the configuration file based on order of appearance, i.e., the marker appearing uppermost in the file is given highest priority.

Additionally, a URL validation procedure for video streaming was required, whereby the web address of each video had to match a predefined location to enable videos to be streamed only from a location specified by Aller Media.

Finally, there was also one performance requirement. To ensure a pleasant user experience, the webcam resolution had to be at least 320 pixels in width and 240 pixels in height. However, a more preferable resolution would be at least 640 and 480 pixels, respectively. Resolution has a major effect on the time it takes for ALVAR Flash Library to process a frame when detecting markers. Thus, an increase in resolution leads to a decrease in the amount of frames per second displayed to the user. The application needed to run at less than 200 milliseconds per frame, which equates to 5 frames per second. This was a reasonable requirement as ALVAR Flash Library is allowed a maximum of 100 milliseconds per frame in order to ensure that the application itself has 100 milliseconds for 3D rendering and other processing.

```
<?xml version="1.0" encoding="utf-8" ?>
<conf>
  <general>
    <!-- Timer delay in milliseconds
    for stopping playback -->
    <timerDelayForStoppingPlayback>
      5000
    </timerDelayForStoppingPlayback>
    <!-- Timer delay in milliseconds for
    changing the currently used marker -->
    <timerDelayForChangingMarker>
      300
    </timerDelayForChangingMarker>
  </general>
  <!-- The list of markers that are
  used by this application -->
  <markers>
    <marker>
      <!-- Marker ID -->
      <id>1</id>
      <!-- The video to be played
      with this marker -->
      <url>video.flv</url>
    </marker>
  </markers>
</conf>
```

Figure 5. Configuration file.

D. Implementation

The implementation of our web-based video streaming application is illustrated in Fig. 6. The core of the application is composed of two ActionScript classes. The main class is for the main application and the ALVAR class is for exchanging data with ALVAR Flash Library. The main class is responsible for loading the configuration file, initializing a webcam and the application in general and processing each frame. The ALVAR class is used only for passing data to and receiving it from ALVAR Flash Library.

The configuration file used by the main class is in XML format. This was chosen because it is a very common format and also because Flash has native support for reading and parsing XML. The structure of the configuration file is intuitive and easily extendable. The basic information containing the two timer delays mentioned above is located under a dedicated XML element, while the marker information is located under another element. The element for the basic information contains an individual element for each information item. Similarly, the element for each marker contains elements for each marker's information item. All marker elements are listed under the marker information element. Additional elements, which can be read by the application, can be easily added to the structure, enabling the application to be extendable with minimal effort.

Pose and position information is received from ALVAR Flash Library as a transformation matrix, which is a 4 by 4 matrix. This format is also used by the 3D plane for representing its pose and position. Thus, the matrix received from ALVAR Flash Library can be directly set as the 3D plane's new transformation matrix without any additional processing.

The transformation matrix is also used when the 3D plane is presented so that it faces the camera at all times (billboarding). In that case, the transformation matrix must be modified before it is applied to the 3D plane.

Streaming a video in Flash is a straightforward process. Flash has good support for initiating and controlling a streaming session, including a set of events for informing of any changes in the stream's status. In addition, the received stream can be set as the 3D plane texture directly without any pre-processing.

The ALVAR class acts as the interface to ALVAR Flash Library, as mentioned above. It is responsible for providing webcam frame data to ALVAR Flash Library, which it does by utilizing shared memory space, as discussed in Section III B. The ALVAR class allocates the required memory space in

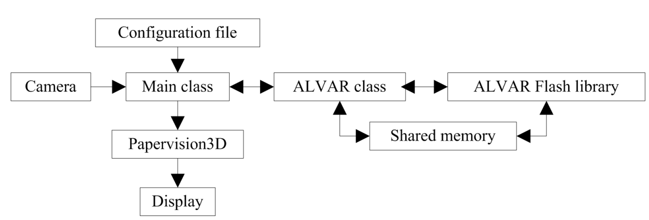


Figure 6. The implementation of web-based video streaming application.

its constructor function, while placing data in the memory space and receiving a response from ALVAR Flash Library for each frame that is handled by another function.

From the end-user perspective, the application is highly intuitive and easy to use. For example, the installation of the application compared to the Dibidogs application is extremely easy. While the Dibidogs application requires to be downloaded from the Internet and installed separately on each computer it is used on, in our application the user is just required to go to a web page where the application resides. If the webcam is not found, the user is merely required connect the webcam and to click a button to re-initiate the application. The user then simply has to hold a marker in front of the webcam in order for the application to detect it and show content on top of it. These tasks are extremely simple and trouble-free for the user, thus enhancing the user experience.

As the application was targeted to be used for marketing, the integration was also carried out in co-operation with Aller Media and the application was installed on their web server. Prior to integration the application was thoroughly tested and final functionality tests were conducted after the installation to verify full operation. The application was then released to the end-users by Aller Media. Fig. 7 shows two pages from Seiska magazine (circulation of 211,000) containing video markers promoting the film Vares – Pahan suudelma.

IV. Testing

The application was tested in order to find out how processing power affects the video streaming and playback performance. This test was conducted with two computer configurations, a low-end and a high-end one. The low-end configuration represented a computer that a common user would have in the near future. Its configuration consisted of Intel Core2 Duo P7450 2.13 GHz processor, 4 gigabytes of RAM and NVIDIA GeForce 9600M GS graphics card. On the contrary, the high-end configuration represented a high-performance computer nowadays rarely available to common users. This configuration contained Intel Core i7-2600K 3.40 GHz processor, 8 gigabytes of RAM and NVIDIA GeForce GTX 570 graphics card. Both configurations had 64-bit Windows 7 and a version 14 of Google Chrome web browser installed, which was used in the tests. The resolution of the used webcam feed was 640 x 480 pixels.

Figure 7. Two pages from Seiska magazine.

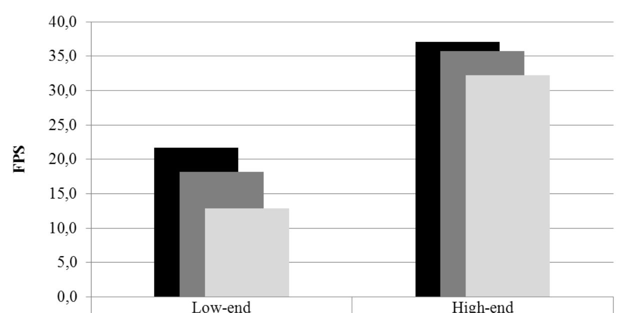
The test set out to establish three performance ranges, namely minimum, maximum and average. Minimum and maximum represent the lowest and the highest performance, whilst average indicates the performance in the long term. The values used to indicate performance were measured within a ten second time span and the unit of measurement was frames per second (FPS). FPS describes the frame rate of the display and it can in turn be used to define the application's usability. Based on our rough estimate with 60 frames per second the usability would be excellent, whilst 20 FPS would still enable very fluent usability. 10 FPS is the limit at which the performance is still acceptable, but with 5 frames per second the frames are updated so slowly that the interactivity with the user and the application suffers too much and the application becomes unusable.

From the results (Fig. 8) it can be seen that the performance is good with both computer configurations. The low-end configuration has a fluent frame rate with an average of 18.2 frames per second, while the high-end configuration's performance is very fluent reaching all the way to 35.7 frames per second. Thus, the high-end configuration is approximately two times faster than the low-end one.

V. Discussion

The Flash video streaming concept was to enable video playback on top of a marker visible in a webcam feed. The application was developed for Aller Media in co-operation with them and the requirements for the application were fully achieved. As Section IV illustrated, the application's performance is good. The FPS value is clearly over 10 frames per second using both computer configurations. The results thus indicate that the application has very good overall performance and that its implementation was successful.

Aller Media was very pleased with the application and especially appreciated the fulfillment of their requirements



	Low-end	High-end
Max	21,7	37,0
Average	18,2	35,7
Min	12,8	32,3

Figure 8. The results of the performance test.

and the readiness of the technology for AR applications. In addition, Aller Media collected feedback from their end-users. At its best, the application had over 360 users per day and the average user used the application for more than 3 minutes. The feedback was mostly positive and the application created considerable excitement. The application also received considerable attention after the customer project with Aller Media, as it was followed by two other commercial projects. In addition, numerous other customers have shown interest in the application.

The application's performance could be improved with version 11 of Adobe Flash Player, released in October 2011. Regrettably, the release of version 11 came after the application had already been designed, implemented and deployed. The new version has support for hardware accelerated graphics rendering, which enables rendering to be executed by the graphics card, producing significant performance gains. However, having the advantage of the new version of Flash Player would require changing the 3D engine used as Papervision3D currently has an inactive development status, its latest version having been published in December 2009.

VI. CONCLUSION

In this paper we presented a web-based augmented reality video streaming application for marketing purposes. The application was commissioned by Aller Media and it was used for promoting two Finnish movies in Seiska magazine. Our solution thus combined conventional print media with digital media. The core functionality of our application was to identify markers from webcam video, to stream a video, and to show the video in a 3D plane on top of the marker in such a way that the video responds to the marker's movement. VTT's augmented reality library ALVAR was converted to Flash in order to use its marker detection capabilities in the web environment. The feedback collected from Aller Media and the end-users showed that the application is well suited for marketing purposes. Aller Media was very pleased with the application and especially appreciated its fulfillment of their requirements and the readiness of the technology for AR applications. The application had over 360 users per day and the average user used it for more than 3 minutes. The test results indicate that the performance of the application is good as the frames per second value sets between 18.2 and 35.7 with low-end computer and high-end computer, respectively. Additionally, the distribution of the application is efficient and thus the application is well suited for the web environment.

REFERENCES

[1] R. T. Azuma, "A Survey of Augmented Reality," Presence: Teleoperators and Virtual Environments, vol. 6, no. 4, Aug. 1997, pp. 355-385.

[2] W. R. Sherman and A. B. Graig, Understanding Virtual Reality: Interface, Application, and Design. Morgan Kaufmann, San Francisco, USA, 2003.

[3] H. Schumann, S. Burtescu, and F. Siering, "Applying Augmented Reality Techniques in the Field of Interactive Collaborative Design," Lecture Notes in Computer Science, vol. 1506, 1998, pp. 290-303.

[4] Paasilinna (accessed on 28 November 2011) Dibitassut ja kadonnut prinsessa. URL: http://www.paasilinna.fi/kirjat/dibitassut-ja-kadonnut-prinsessa.

[5] T. Harviainen, O. Korkalo, and C. Woodward, "Camera-based Interactions for Augmented Reality," Proceedings of the 5th International Conference on Advances in Computer Entertainment Technology, Athens, Greece, Oct. 29-31, 2009.