

# Cultural Heritage Layers: Integrating Historic Media in Augmented Reality

Michael Zoellner, Jens Keil, Timm Drevensek and Harald Wuest  
Fraunhofer IGD

**Abstract**—In this paper we are presenting Cultural Heritage Layers, an approach that enables the visualization of historic media like drawings, paintings and photographs of buildings and historic scenes seamlessly superimposed on reality via video see through using X3D. This enables simple, inexpensive and sustainable Augmented Reality applications in the cultural heritage and architectural area based on industry standards.

The main idea is to use existing historic media from archives and superimpose them seamlessly on reality at the right spot. These locative layers are context sensitively telling the location's history and create the impression of a virtual time journey.

The registration of the virtual objects in the video images is provided by a robust 6DOF tracking framework based on two technologies that work in tandem: an initialization step based on Randomized Trees and a frame-to-frame tracking phase based on KLT.

The entire application runs in real time on current Ultra Mobile PCs and MIDs.

**Index Terms**—Capture Technology and Delivery Platforms, Mixed Reality, Computer Vision for Cultural Heritage, Enhanced Environments and Locative Media

## I. INTRODUCTION

Over the last years we observed four major problems of Augmented Reality in cultural heritage and tourism projects: 1. Virtual reconstructions of buildings often suffer from poor visual quality, doubtful scientific accuracy and at the same time their produce is somehow expensive. 2. Robust markerless tracking solutions are often missing, which leads to the use of non aesthetic markers. 3. There are still no high-performance and portable computing platforms commonly available to run demanding software solutions. 4. Custom software solutions are often based on proprietary software and are thus not long term sustainable.

In this paper we are presenting solutions for the first three problems based on our experience in cultural heritage projects. Furthermore we are proposing X3D as a platform for sustainable cultural heritage applications based on rich media like Augmented Reality.

Most interactive cultural heritage projects suffer from a lack of 3D content for interactive visualizations. The reason is not only the expense of the creation of high quality 3D models or the usage of laser scanning technologies. Scientific accuracy is another serious matter. Untrue photorealistic interpretations of historic objects can lead to be fact in the viewers mind. Furthermore exaggerated and wrong lighted 3D models are ruining every Augmented Reality scene. Thus we have developed a way for the intelligent use of existing historical content like drawings and photographs to enable more cultural heritage



Fig. 1. Video see through effect with Reality Filtering on UMPC

sites to use Augmented Reality to enrich the experience of their sites. Instead of rendering 3D models on the real view we are rendering drawings and photographs, originally taken from the views position, right at the place. Additionally the video stream of the reality is rendered to fit the source material, for example black and white drawings or photographs, oil paintings or aged video content.

Another difficulty of using Augmented Reality in cultural heritage projects is that the tracking technology has to be robust enough to handle large environments with difficult lighting conditions, and must be at the same time discreet enough to not disturb the user, hence the tracking process should be invisible to the user, excluding the usage of standard markers. In this application, we opted for a mixture of two recent markerless tracking technologies, which proves to be robust and fast enough for our aimed platform. The usage of 2D textures instead of 3D models also needs less computing power and thus enables the visualization on ultra mobile PCs like the Sony UX and mobile phones in the near future.

The remainder of the paper is organized as follows. Section II is a global overview of our system. Section II-B presents briefly the markerless tracking technologies we are using for outdoor scenarios. We present in section III applications of our approach at two cultural heritage sites. Finally we conclude in section IV and suggest future work.

## II. SYSTEM OVERVIEW

### A. Hardware setup

Our system adresses visitors of archeological, historical sites or museums. The equipment consists of a hand-held Ultra Mobile PC (UMPC), on which our software is running. The UMPC serves as a *reality-window*: a small, integrated video camera placed on the back of the device records live pictures that are shown in full screen on the 5-inches display. Thus, the user sees the reality through the device (see Figure 1).

For a given application site (*e. g.* a museum or any of those mentioned above), a number of objects of interest termed *spots* are selected by the application designer. These spots can be for example sculptures or paintings, building facades, temple ruins, or even gardens. When looking through the UMPC, these spots will be *augmented* with additional information in form of original historical content like drawings or paintings of buildings and historic scenes.

### B. Tracking system for outdoor scenarios

Given a few reference images of the spots, our system learns in a preliminary phase to recognize the spots. In the online phase (on site), the system is first in a *initialization* routine, which analyses the images from the camera and finds out if the current view is from a given spot or not. When a spot has been recognized, the system enters a *tracking* state, where the position of the object of interest is precisely recovered and tracked, allowing for a seamless geometric integration of the additional information. These two phases (initialization + tracking) take advantage of recent advance in computer vision in a novel multiple object detection and tracking framework, and are described in more detail in our previous work [3].

We are concentrating on one main scenario naturally arising at cultural heritage sites: distant objects like buildings or landscapes. In this case, the size of the object is small in comparison with the distance from the user, and the objects can be approximated as a panoramic view. In our system, we developed a texture based tracking algorithm, which relies on reference images of the observed scenario to estimate the camera pose.

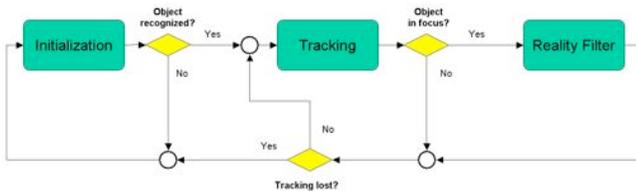


Fig. 2. Block diagram of our multiple scene augmentation algorithm

Figure 2 shows a block diagram of the logics of our system. The system first enters an initialization state, where an object/scene detector searches for known spots, based on feature marks which were trained during the system’s learning process. When an object has been discovered, the position and orientation of the object is computed. Then, the system enters the tracking phase, where the position of the object is updated in a frame-to-frame basis. At the same time, the system looks if the object of interest is in the focus (a predefined zone in the field of view, roughly in the middle of the image). If so, the real images are processed by the Reality Filter and the original content is shown on the screen.

Given a few reference images of the spots, our system learns in a preliminary offline phase to recognize the spots. In the online phase (on site), the system is first in an *initialization* routine, which analyses the images from the camera and finds

out if the current view is from a given spot or not. Therefore a feature point classification approach based on Randomized Trees [1] is used. For every reference image a set of trees for well detectable key-points is created in the pre-processing phase. The initialization module tries to detect the key-points of only one reference image in every frame. This strategy limits the computational costs and guarantees a constant frame rate during the initialization step. When a reference image has been recognized, the system enters a *tracking* state, where the position of the object of interest is precisely recovered and tracked, allowing for a seamless geometric integration of the additional information. The tracking module is based on the alignment of patches of a reference image. At distinct points patches are extracted in the reference images and located with the KLT-tracking method [2] in the live video stream. Since this tracking algorithm is a local search method, it always needs a rough prediction of the feature positions. However, the feature points can be tracked with sub-pixel accuracy, and more stable results for the camera pose estimation can be produced than with purely using the approach based on randomized trees. Our method with these two phases (initialization + tracking) takes advantage of the benefits of both Randomized Trees and the KLT-tracker. More detail can be found in [3].

In order to let application developers and designers create the tracking for a scene we created a simple automation in our software. The user only selects a reference image for the tree generation. In our example of Temple Diana we used tourist’s images of Reggia Venaria Reale found on the internet for tracking. The automation creates the trees and an example X3D scene with the tracking.

There are two advantages of this system compared to prior Augmented Reality content creation and tracking. First we are able to use a printed poster for testing the tracking and the placement of the content. Even whole rooms can be tested with cardboard miniatures. Thus traveling time and costs for on-site testing are reduced and the stability of the application increased. Once arrived at the site the application runs out of the box. Our application of Reggia Venaria Reale trained with sunny tourist photos even worked on a rainy April day with desaturated colors. Only some modifications were needed for a first presentation.

### C. X3D scene

In correspondence to the available hardware on Ultra Mobile Computers, we are presenting a scene setup that only needs minimal OpenGL and 3D acceleration. Therefore we are only using 2D textures of historic buildings and scenes instead of reconstructed 3D models. An X3D scene for this scenario has the following structure: A camera image in the background is enabling the video see through effect. A fixed, background filling quad is used to display the video footage. Therefore we have extended the X3D Standard [4] with a pixel based EnvironmentNode called *PolygonBackground*. In front of the background overlays showing a building, painting or facade is located as a flat ImageTexture. On top of this layer all interactive content is placed. These points of interest consist



Fig. 3. Left picture: Overlay of the original Reichstag of 1901. Right picture: Stacked urban grain plans on satellite image of Berlin

of media like images, movies and text through standard X3D nodes like ImageTexture, MovieTexture and Text. Interactivity is achieved with standard X3D PointingSensors and custom extensions like UserBody [5], [6]. The UserBody nodes are ray-intersect or collision triggers for the high-level PointingDeviceSensor instances. Interactive points of interest are activated either by a roll over effect triggered by a cross hair like UserBody in the center of the display or by a TouchSensor when touching it on the screen.

Upon the described extension and a heavily useage of Followers for smooth motion, we have created several Protos to reduce the X3D Interface to a minimum. A collection of this are:

- The *BubbleOverlay*-Proto allows to extend a real object with appendable information by clicking onto augmented marks. Sticked informations are getting faded in and will blind out when the point of interest is getting out of sight.
- The *TimelineOverlay*-Proto acts like a View-Master-Device [7], where the user can switch between several versions of the same object. This allows switching between the variations of different historical moments in time (Figure 3).
- Finally the *LayeredOverlay*-Proto overlaps multiple transparent images with a small offset. The resulting effect is a simulation of depth for a straight lined 3D object without the cost of rendering many triangles (Figure 3).

These prototypes reduce authoring of a scene to a minimum. For example a scenario with 5 points of interest on a building takes only 5 lines of X3D code within the Scene node.

### III. APPLICATIONS

The two application scenarios, which we are presenting, are basically working like described in II-C. While interactivity technically is achieved with X3D Pointing sensors, the interaction metaphors itself are application specific and adopted to fit best in the scenarios.

#### A. 20 Years since the Fall of the Berlin Wall

In this application we revived the long and rich history of Berlin: After the Second World War the entire city was almost completely destroyed and during 1961 until 1989 the city centre was divided by a huge wall. 1933 the Reichstag, one of Berlin's most famous sights and now domicile of the German Parliament, was burned down and its cupola destroyed.

To gain tourists better access to the historical photographs of this time and to the changing architecture of the building, we used again the Sony UX's display for video see through in an outdoor situation. Photographs from different times are seamlessly superimposed on the real scene, showing and explaining development of landmarks and history (Figure 3 left). The interface is kept very simple. By fading in information sensitively to context and location. Touching a virtual overlay on the display, the media data traverses visually from left to right through the time: The content is not switched abruptly but transitioned smooth and comprehensible from one picture to another. This visuall effect was chosen to provide a better usability and perception and finally to increase immersion and familiaristion with this technology, which is still laking in today's Augmented Reality applications. Anyway, these techniques are well known and oftenly used in webpage user interfaces based on Adobe's flash platform or todays JavaScript/AJAX implementations.

Indoor, in a museum the center of the project is a table with a satellite image of Berlin on which a 3D model of the Berlin Wall and the urban development from 1940 - 2008 are displayed. Therefore urban grain plans showing areas covered with buildings are augmented on the satellite image (Figure 3 right).

While the visualisation of thousands of buildings would be very cost intensive. In addition to the described tracking technique the limited hardware of the UMPC would be overstrained. To create the aspect of augmentation in this example it is important to create the look of depth. We did this by layering multiple layers of the same image of the grain plan over each other with a little offset. While each image contains an alpha channel the final projection leads to a pitted but 3D looking visual experience without the need of much hardware acceleration. For a better authoring of the X3D scene this functionality was encapsulated in an own X3D prototype node like described in section II-C.

#### B. Reggia Venaria Reale

This approach is already used for the field tests of the EU funded project iTACITUS [8]. Within this project, one of the field test areas of the Augmented Reality applications is Reggia Venaria Reale, an UNESCO World Heritage site in Italy close to Turin. The former residences of the Royal House is comparable to the french Versailles. The site has been restored over the last years and was opened to the public in fall 2007. While there are only a few 3D reconstructions of some buildings there is a vast archive of historic drawings and paintings. There are frontal drawings of fasades of complete streets and the main palace's buildings.

This vast archive is the basis of our visual time machine. With the Reality Filters we are able to create this effect through the display of a Sony UX Ultra Mobile PC. In order to match the source material we are rendering the video stream like a black and white drawing (Figure 4). Even visitors, guides and guards are rendered in black and white, thus granting a much more realistic and authentic look and feel (without

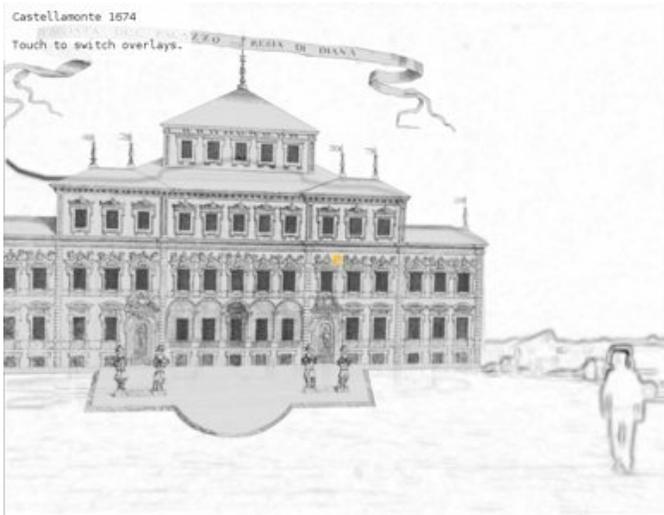


Fig. 4. Reality Filter enables seamless integration and removes construction site

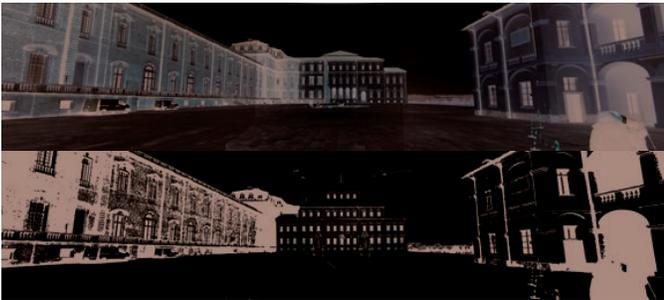


Fig. 5. Palazzo di Diana's architecture over the centuries.

higher rendering techniques), where augmented media doesn't look like foreign objects. Finally, the whole scene looks like a real time ancient drawing.

This Reality Filtering [3] approach is a solution to adapt the real view of a scene to blend well with black and white drawings in Augmented Reality applications by applying an inverted sobel filter on the video stream. In contrast to the methods described in [9] we are concentrating on 2D textures instead of 3D models.

Reggia Venaria Reale's Palazzo di Diana's architecture was modified several times over the years. Each state of the buildings was documented on drawings. We are superimposing these drawings of the modifications of the architects Castellamonte (1674) and Garove (1700-1713) on the facade of the main building. Visitors standing in the large courtyard are seeing the fountain and the current restoration of the palazzo rendered like a drawing. While listening to the audio guide's story about the palazzo the buildings appearance is switching through the centuries while seamlessly integrated into the environment (Figure 5).

#### IV. CONCLUSION

In this paper, we have presented our Augmented Reality concept called Cultural Heritage Layers. Its main features

are the possibility to show original 2D content superimposed on lookalike reality, its robust markerless tracking and its portability on small devices like UMPCs and MIDs. The applications are written in X3D and rendered by the InstantReality framework [10].

The adaptation of the reality is made possible by the use of interchangeable filters that enable a better integration of the ancient content in the reality. The tracking technology is based on a fast initialization followed by a robust 2D tracking of local features. This enables the suppression of planar markers usually seen in Augmented Reality scenarios.

The system can be used with a simplified camera model where only the camera's rotation is computed. We showed results of our application in the area of cultural heritage, where the system runs on an Ultra Mobile PC (Sony Vaio UX) with 15 frames/sec. For now the Reality Filters and the 2D overlays can be chosen by the application developer or online by the user. In a future version of our system, we will investigate the automatic detection of the right filter for the best integration of the content in the real image.

#### REFERENCES

- [1] V. Lepetit, P. Lagger, and P. Fua, "Randomized trees for real-time keypoint recognition," *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 775–781, 2005.
- [2] C. Tomasi and T. Kanade, "Detection and tracking of point features," *Carnegie Mellon University Technical Report CMU-CS-91-132*, Apr. 1991.
- [3] M. Zoellner, A. Pagani, Y. Pastarmov, H. Wuest, and D. Stricker, "Reality filtering: A visual time machine in augmented reality," *VAST 2008. Proceedings*, December 2008.
- [4] Web3DConsortium, *X3D*, 2008, <http://www.web3d.org/x3d/specifications/ISO-IEC-FDIS-19775-1.2/index.html>.
- [5] J. Behr, P. Dähne, and M. Roth, "Utilizing x3d for immersive environments," in *Web3D '04: Proc. of the ninth int. conf. on 3D Web technology*, Fraunhofer IGD. NY, USA: ACM Press, 2004, pp. 71–78.
- [6] J. Behr, P. Dähne, Y. Jung, and S. Webel, "Beyond the web browser - x3d and immersive vr," in *IEEE Virtual Reality 2007: Symposium on 3D User Interfaces (3DUI)*, Fraunhofer IGD. USA: IEEE, 2007.
- [7] Sawyer'sPhotoServices, "View-master," 1939, <http://en.wikipedia.org/wiki/Viewmaster>.
- [8] iTACITUS, "Intelligent tourism and cultural information through ubiquitous services," 2008, <http://www.itacitus.org>.
- [9] J. Fischer, "Stylized augmented reality for improved immersion," *Proceedings of IEEE Virtual Reality*, pp. 195–202, 2005.
- [10] Avalon, "Avalon," 2008, <http://www.instantreality.org>.

Citation:

Zöllner, Michael ; Keil, Jens ; Drevensek, Timm ; Wuest, Harald:

**Cultural Heritage Layers: Integrating Historic Media in Augmented Reality**

In: Sablatnig, Robert (Ed.) u.a. ; International Society on Virtual Systems and MultiMedia:

15th International Conference on Virtual Systems and Multimedia. Proceedings : VSMM 2009.

Los Alamitos, Calif. : IEEE Computer Society, 2009, pp. 193-196.