FROG

FUN ROBOTIC OUTDOOR GUIDE

Deliverable: D2.3

Sensor data and Multimedia database with the virtual content for AR application

Consortium

UNIVERSITEIT VAN AMSTERDAM (UVA)

IDMIND - ENGENHARIA DE SISTEMAS LDA (IDM)

UNIVERSIDAD PABLO DE OLAVIDE (UPO)

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE (ICL)

UNIVERSITEIT TWENTE (UT)

Grant agreement no. **288235**

Funding scheme **STREP**

SEVENTH FRAMEWORK PROGRAMME

DOCUMENT INFORMATION

Project

Project acronym:	FROG
Project full title:	Fun Robotic Outdoor Guide
Grant agreement no.:	288235
Funding scheme:	STREP
Project start date:	1 October 2011
Project duration:	36 Months
Call topic:	ICT-2011.2.1 Cognitive Systems and Robotics (a), (d)
Project web-site:	www.frogrobot.eu

Document

Deliverable number:	D2.3
Deliverable title:	Sensor data and Multimedia database with the virtual content for AR application
Due date of deliverable:	M36 - September 2014
Actual submission date:	M36+ - October 2014
Editors:	UT
Authors:	UT
Reviewers:	IDM
Participating beneficiaries:	UT
Work Package no.:	2
Work Package title:	Robot Control, Navigation and Location Based Content
Work Package leader:	UPO
Work Package participants:	YD, IDM, UPO, Imperial, UT
Estimated person-months for deliverable:	2
Dissemination level:	Public
Nature:	Other
Version:	Revision
Draft/Final:	Final
No of pages (including cover):	10
Keywords:	Sensor data, Multimedia content, AR application

Contents

1. Introduction

This deliverable is a revision to replace deliverable D2.3 from May 2013 written by YDreams. In this revision we will present how sensor data is used as a trigger to display or render content of the FROG tour. We use the position and the orientation for the FROG to play different kinds of multimedia content. Position and orientation data can be a trigger for playing multimedia content or it can be used to properly render the content in the AR application (e.g. to overlay a 3D model in the camera view at a location). This application is used for the scripted full FROG tour in the Royal Alcázar.

We will explain this approach on the basis of a map of the Royal Alcázar. This map is the one that is used by UPO's localization module, resulting from the captured laser range measures. This map has been annotated with information to display multimedia content inside the editor of the Front End application (see Figure 1).

In Section 2 we will shortly discuss the sensor data we are using to display content. Section 3 will discuss how this data is used to display multimedia content.

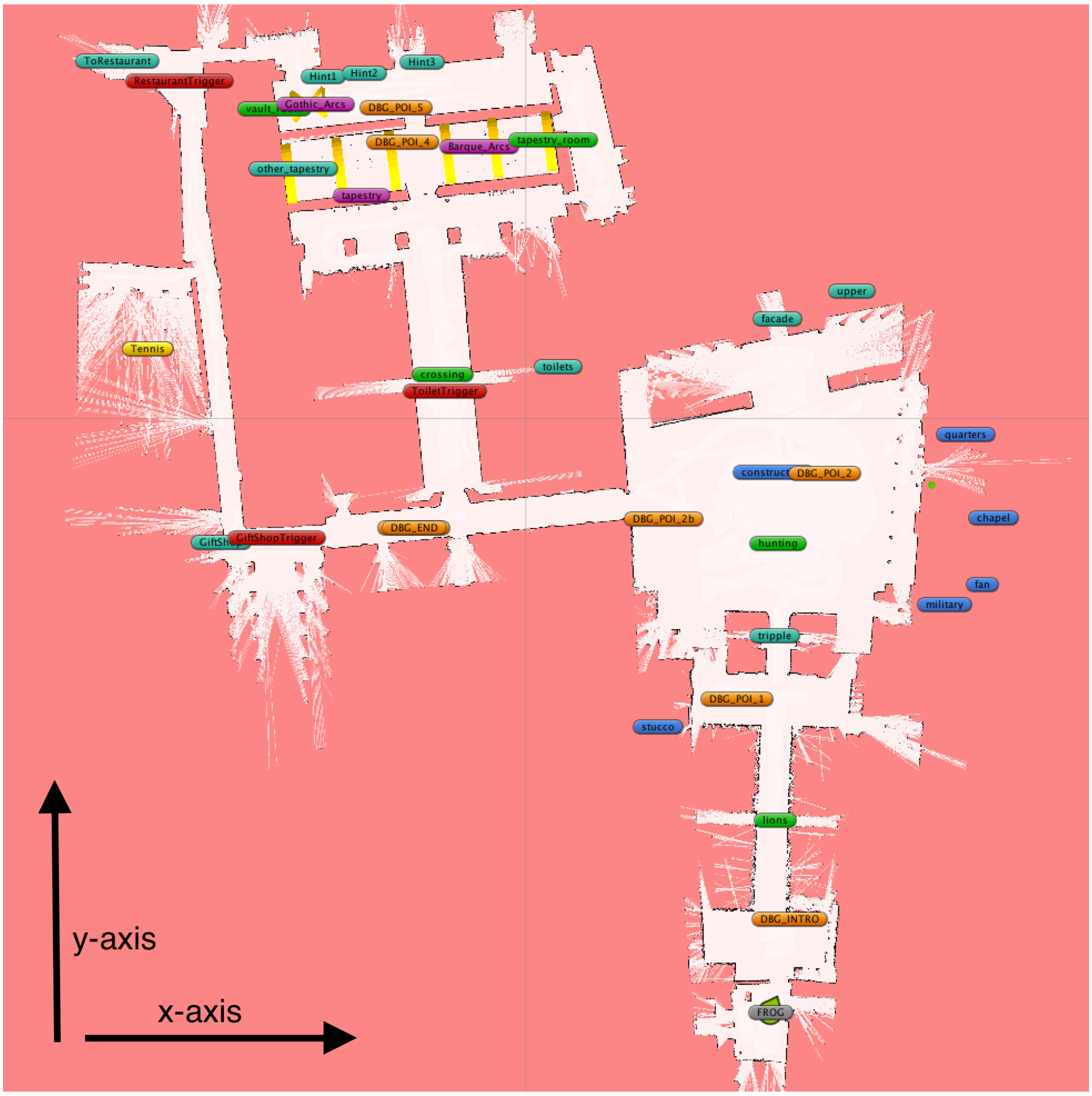


Figure 1 Map of the Royal Alcázar showing different kinds of multimedia content

2. Sensor data

We use the editor of the game-engine Unity3D to create the Front End application. The UPO map is used as a floor plan of the tour's "level". Multimedia content is placed meaningfully into the map. A true-to-scale model of the FROG roams this level.

To achieve real-time congruency between the FROG's location in the virtual level and the actual location in the Alcazar, we use the location (x, y) and orientation (around the z-axis) of the `\FROG_pose` topic provided by the UPO navigation module via the RosBridge. FROG pose is received and processed by the State Machine and forwarded to the Front End. The position of the FROG is drawn in Figure 1 (the grey label "FROG" at the bottom of Figure 1).

This information is sufficient to make use of trigger zones in the level that results in specific robot behaviour to be executed – such as changing volume in certain areas, or making location remarks (see Section 3.5), but for presenting real-time camera images with superimposed Augmented Reality content on the FROG screen, we further require the extrinsic parameters of the camera – namely position and orientation in reference to the world.

We rely purely on location data from the navigation component and the sensors inside the servomotors of the arm, rather than attempting a vision based approach to derive camera extrinsic. For this, a true-to-scale model of the FROG was extended with a model of the arm (see Figure 2) in which the camera used for AR purposes is mounted. The angular positions of the servomotors in the three joints are read out and inform the angular position of their virtual counterparts. In Section 3.1, we describe in more detail how this model and the feedback of the servomotors serves to realize the Augmented Reality effect.

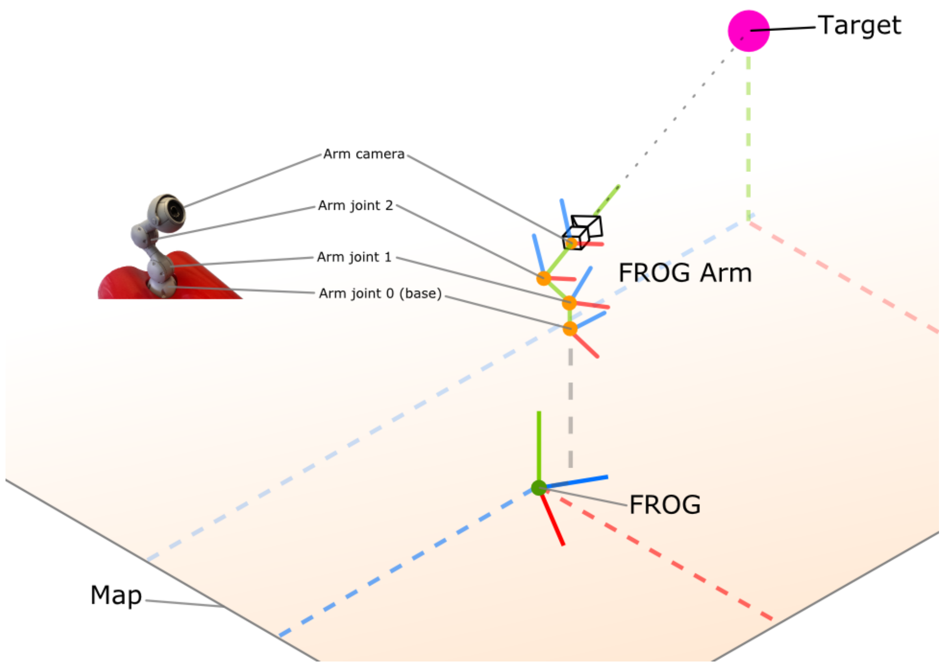


Figure 2 Model of FROG (including a model of the arm on top) inside the navigation map - used as model of "the world". We may place content inside this map and have the arm "look at" the content using camera targets

3. Multimedia content

The position of the FROG is used to trigger or render different of types ofmultimedia content to be displayed. In this section we will present an overview of different kinds of media that can be triggered or rendered based on the position and orientation of the FROG. This multimedia content is part of the Behaviour Library that is discussed in deliverable D2.4.

3.1 Camera target

To incorporate camera images of constructions into the tour and to render additional content such as images and 3D models properly over those camera images, the application does not only need to know where the camera is currently located (as explained in Section 2), but also how to manipulate the FROG – or in particular: FROG's arm, in which the AR camera is mounted – in order to achieve a certain camera view of the location.

In Figure 3 we illustrate the position and orientation of a model of FROG inside our map of the Alcázar (Figure 1). As already mentioned, this position and orientation is determined by the navigation component. We further have available the aforementioned model of FROG's arm, with the three joints and the camera mounted inside the end of the arm.

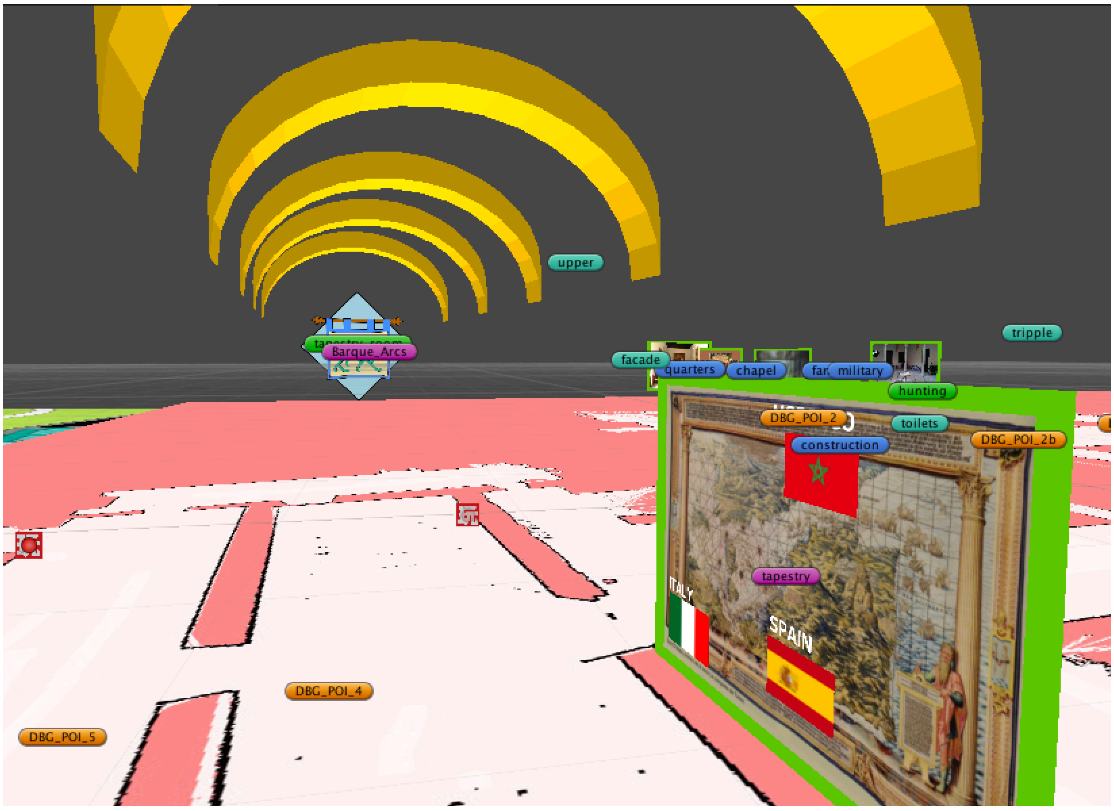


Figure 3 Multimedia content on the map of the Royal Alcázar; 3D model of the arcs, images and camera targets

This model provides sufficient information[1] to calculate, using inverse kinematics, the angle for each joint to have the arm's camera displaying the desired target in the centre of its view. Figure 4 presents an example of two camera targets on the map and the corresponding targets in the real world.

To produce the Augmented Reality effect, a virtual camera is located inside a copy of the model – with the same coordinates and orientation of the real camera. This copy of the model is driven in real time by the actual readouts of the servomotors, as mentioned in Section 2.

This virtual camera renders all "AR" marked content inside the level - matching also the intrinsic parameters (such as field of view) of the virtual and real camera allows us to superimpose the virtual renderings onto the camera image to achieve the Augmented Reality effect.

The different virtual objects that are described in detail in the following sections can be placed the same way we place camera targets - on the virtual representation of the map.

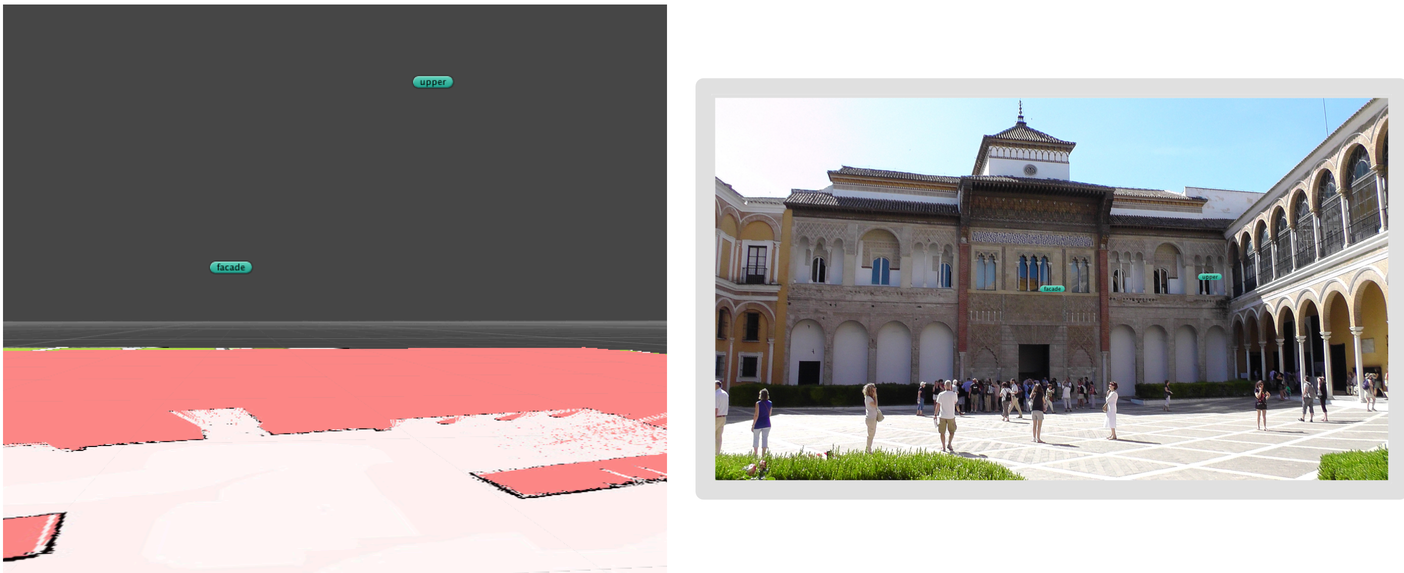


Figure 4 Example of two camera targets. On the left side we see the camera targets in the map, on the right side we see the corresponding targets in the real world

3.2 3D model

3D models can be displayed on FROG's touch screen. Based on the position and the orientation, and by looking at the 3D model with the arm camera on the map, the 3D model will be visible in the scene. 3D models can be rendered as overlays on top of real world objects (e.g. Figure 5 or they can be renders in space (e.g. Figure 6)

[1] The results could be improved by also incorporating the FROG's rotations over the x- and y-axis, as measured by the accelerometer module.

Figure 5 Overlay of a 3D model on top of real world objects at the in the Tapestry room

Figure 6 3D models of the Gothic arches. On the left image the model on the maps at Point of Interest 5. On the right image the model as an overlay over the camera view shown on the touch screen

3.3 Images

Images can be displayed as multimedia content. Images are placed and stored on the map. By looking into the direction of the image with the arm camera the image will be visible in the scene in real time. Figure 7 presents an example of an image placed on the map.

Video content is supported in the same way as images. Video content was used in the June 2014 demonstrator, but is not part of the demonstrator since the September 2014 update. Video content was less flexible and was replaced by separate audio clips and associated images.

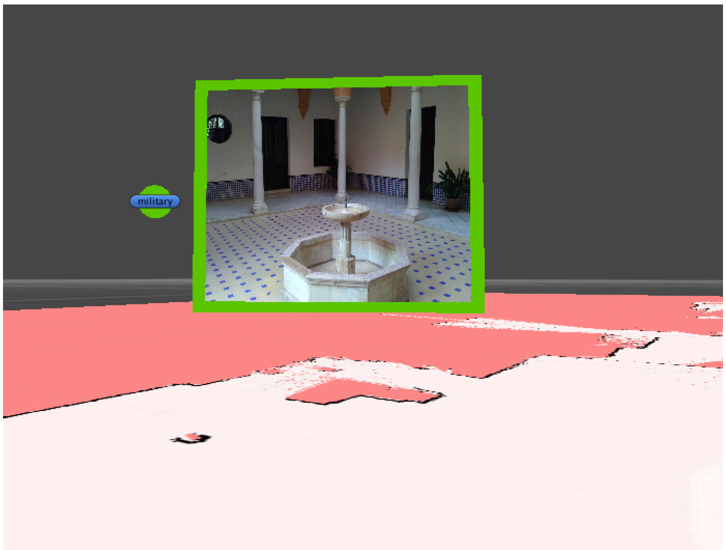


Figure 7 An example of an image as part of the content

3.4 Audio clips

Audio clips can be triggered based on the position of the FROG. Figure 8 gives an example of an audio clip that will start fading-in the sound of a tennis match when the robot drives by the Jardin de la Alcubilla, which was a tennis court at the beginning of the 20th Century.

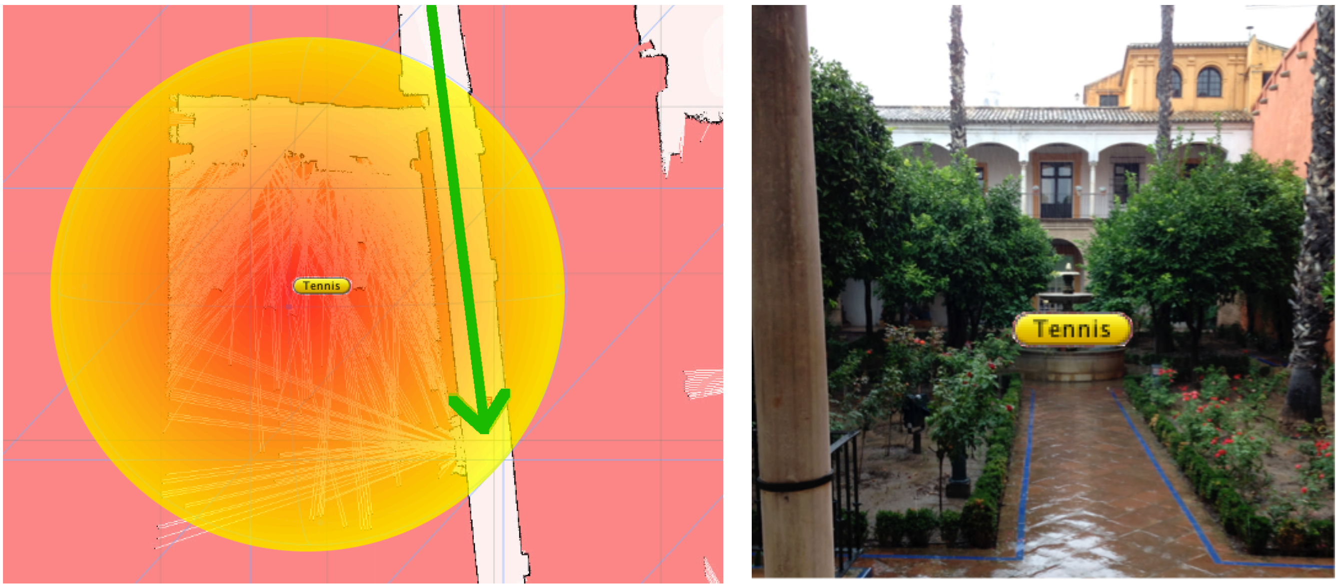
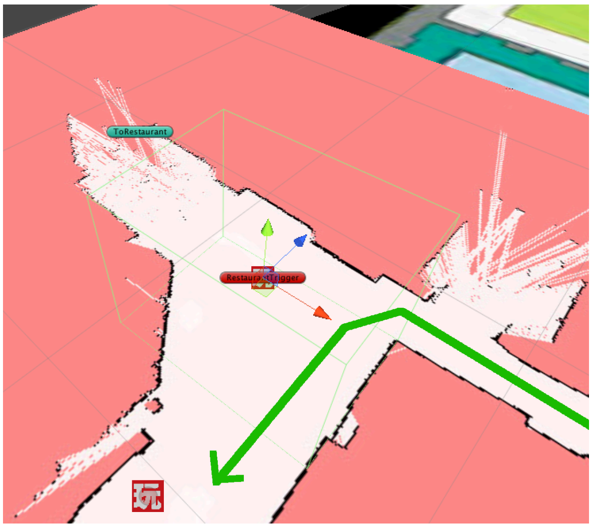


Figure 8 On the left side, an example of an audio clip that starts playing based on the location of the FROG. The closer the FROG is to the midway point of the garden, the louder the volume of the audio clip will be. The green arrow is FROG's path. On the right side, a view on the Jardin de la Alcubilla.

3.5 Facility and event information

Based on the location of the FROG, content about useful facilities nearby can be displayed to the user. These facilities are marked on the map and when FROG drives by these places, it will display information about them to inform the users about the place nearby. Current examples of these places are the toilets, the restaurant and the gift shop. FROG will use the arm to point to the location and an audio clip will tell about the useful facility nearby (e.g. “*Over there you can find the restaurant*” while using the arm to point in the direction of the restaurant as described in Figure 9).



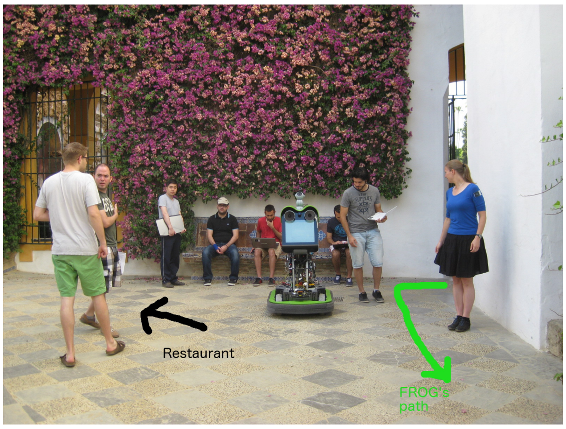


Figure 9 On the left side, an area on the map indicated by the green cube to show where content will inform users about where to find the toilets, restaurant and gift shop (in this case the restaurant). On the right side the position in the real world.

4. Conclusion

We have tried to illustrate how a careful annotation of the navigation map of the mission site (in this case the Royal Alcázar) is used to trigger or render multimedia content. FROG uses its location at any given moment to trigger relevant presentations. Presentations can be altered as a reaction of a location shift (e.g. location based volume control).

While our principal approach is promising, it leaves room for improvements - as is evident by the screenshots provided in this document. Specifically the non-vision based approach to camera-extrinsic estimation is imperfect. Due to the prototypical implementation of these concepts, the source and magnitude of errors have not been investigated thoroughly thus far. Some error also stems from missing information – for one, we do only consider the robot rotation around the z-axis, although on uneven ground, other rotations will also apply.
However, we realize that even with more data, some error will always accumulate between the creation of the floor plan and virtual model of the robot, the placing of the content in the editor and finally, the sensor data received at runtime. However, since we reduce search space of the problem significantly with our first “estimation” of the camera extrinsic, an otherwise too expensive vision-based approach might be a good complement for the “last mile”.