IMMERSIVE PRODUCTION AND DELIVERY OF INTERACTIVE 3D CONTENT

DELIVERABLE D6.1:

DETAILED SPECIFICATION OF USE CASE SCENARIOS

Contractual Date of Delivery:          August 31, 2012
Actual Date of Delivery :              September 5, 2012

Work Package:                          WP6
Dissemination Level:                   Public
Nature of Deliverable:                 Report

Lead Contractor:                       BBC
Coordinating Person:                   Oliver Grau
                                       Oliver.Grau@bbc.co.uk

Project Leader:                        BBC
Work Package Leader:                   ART
Authors:                               O. Grau (BBC) , D. Knossow (ART),  E. Maggio (OMG),
                                       J.S Franco (INRIA), D. Schneider (HHI), A.Hilton (UoS)

# Classification and Approval

Classification: Public

This document has the status 'Public' and may be published as a whole, including this page. Publication of parts of the document must be in unanimous agreement within the RE@CT Steering Board and subsequent EC approval/agreement.

Disclaimer

Neither the RE@CT Consortium nor any of its officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the RE@CT Consortium nor any of its officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage, personal injury or death, caused by or arising from any information, advice or inaccuracy or omission herein.

Acknowledgements

All partners of the RE@CT project contributed to this report.

# RE@CT Consortium Overview

| Participant no. * | Participant organisation name | Short name | Country |
|---|---|---|---|
| 1 (Coordinator) | British Broadcasting Corporation | BBC | UK |
| 2 | Fraunhofer HHI | HHI | Germany |
| 3 | INRIA | INRIA | France |
| 4 | University of Surrey | Surrey | UK |
| 5 | Artefacto | ART | France |
| 6 | OMG | OMG | UK |

# Abstract

This document is part of work package WP6 "Demonstrator and Evaluation" of the RE@CT project. It describes use case scenarios for the technology developed by the RE@CT project, focussed on two cases which will be developed into demonstrators: the first scenario is set within cultural heritage VR[i] or AR[ii] applications in a public environment, typically a museum or exhibition. The second case is an application alongside a TV production, either as a live scenario during the TV production or an online application during and after the broadcast.

The document also describes RE@CT's specific production requirements and a basic layout of the software components that have been already developed or will be during the project, as outlined in the 'description of work'. This specification is more 'production centric' as it describes the production flow and interfaces to the RE@CT components for the professional user of the system. It therefore complements the more technical interface specification of D5.1.1 and D5.1.2.

---

[i] Virtual Reality

[ii] Augmented Reality

# Table of Contents

# 1   Introduction

The RE@CT project is developing a new production methodology and toolset to create film-quality interactive characters from 3D video capture of actor performance. The key innovation is the development of methods for the analysis and representation of 3D video to allow re-use for real-time interactive animation. This will enable the efficient authoring of interactive characters with video quality appearance and motion.

The objective of WP6 is to demonstrate and evaluate the developed tools and methodology based on two finished products that involve the production facilities being developed. The tools should allow the production of finished products that show a fully realistic display of interactive characters, whilst allowing this interactive content to be produced alongside a TV or similar production. Therefore, two different production scenarios will be tested:

> - Development of an interactive application available in a public place, related to serious gaming
> - A TV-production where interactive content is produced in parallel to the recording and exploited in an interactive network application related to the production

This document describes the applications that will be set up to demonstrate the different aspects of the project. Rather than focusing on specific scenarios, this document will describe use case scenarios and will address all the pre-requirements, hardware set-up and software developments that will be required to achieve the demonstrators.

This document is organized as follows: first, we will describe the different expected use cases to demonstrate the strength of the video production paradigm set out in the RE@CT project. For these use cases, we will then focus on the production requirements regarding the hardware and technical set-up. Then, we will present a design and software layout that will be developed along the different production phases and for the final demonstrator.

# 2   Specification of use case scenarios

## 2.1 An interactive application for use in a public space

There is a trend in recent innovative games to interact with the real world. Being in front of a screen, at home, becomes less fun than playing in an outdoor environment or an historical site.

Smartphone and tablet PCs largely contribute to this new trend and more conceptually to this new behaviour. Mobility now becomes part of the game. The increasing graphical and computing power of these mobile devices allows amazing games to appear on these platforms. Game developers understand this and conferences like GDC2012 or SIGGRAPH2012 have confirmed this trend. However, most of the games recently released still remain "on screen". Only a few titles, a minority in practice, are starting to take advantage of mobility. Cameras, GPS, and inertial sensors are not yet widely used.

With WP6, we propose to set up a demonstrator related to cultural heritage, a domain in which project partner Artefacto has experience.

Recently Artefacto canvassed several cultural heritage centres and proposed applications based on augmented reality which would take advantage of the user being on site and interacting with the surrounding environment.  Surprisingly, these proposals were met with a degree of reluctance by some professionals in the cultural heritage sector. The reasons for their reluctance were firstly the current lack of realistic and yet affordable avatar and 3D environment modelling, and secondly the

current lack of applications allowing interaction with the surrounding environment. The RE@CT project addresses some of these problems and we propose to build a demonstrator proving this. In the following paragraph, we will describe a general use case scenario for an application dedicated to cultural heritage where realism is required and where we strongly encourage people to "play and learn" in outdoor environments.

The main idea here is to allow the general public to learn while playing not only at home or in front of a screen but also when visiting the place where the story happens.

### From playing at home to life-size interactive games

One of the key concepts we propose as a demonstrator is to develop a game based on real historical facts. This game can be played as a "standard" game at home but also be played in an innovative manner at the historical location (for instance in a castle).

The main idea is to allow families to play at home with a board game and cards using the latest augmented reality technologies to animate and illustrate the different moves from the players. A camera is supervising the board and when players move cards, things happen on the computer (tablet, possibly) screen. For instance, when a player puts down a warrior card, a warrior appears in augmented reality. When the player moves their card, the warrior walks, runs or may jump. If the player puts down a battle card, the warrior takes his (or her) sword and prepares to fight. This concept is already being tested and explored by different game developers and has proved the attractiveness that new technologies can bring to traditional games. But the main innovative aspect is that the game can also be played on a life-size scale.

Here we enter a new dimension where the game surrounds the players. The players are in the game and living the adventure from inside it. In the case of a history-inspired game, players would be able to play on site, with real castles, forests, battlefields, etc. becoming their playing environment. Players would control their own characters through the use of a mobile platform such as a smartphone or a tablet PC. Thus, if a player walks, their "real" avatar precedes them. If they need to jump, they swipe their fingers and their avatar jumps; similarly they could make their character carry out other moves such as kneeling.  When fighters appear, players can choose their weapon and fight. If they feel hungry, they can visit the kitchen and find out how food was prepared in those days, if they need a new weapon, they can go to the armoury and pick one, and so on.

The idea here is to encourage the general public to visit historic buildings, such as castles by proposing innovative ways to learn about history. As part of the RE@CT project demonstrators, we plan to focus on the history of Brittany during King Arthur's reign.

### From the game concept to the technology

The key concept of the demonstrator already exists and Artefacto has often met with potential customers. However, they always raise the same reservation, in that the 3D characters are modelled and do not appear realistic. 3D avatars are generally dedicated to "on-screen" games and a lack of realism is not problematic.  However, if we are considering learning and playing a life-size game on site, the models and the representation of the models in the game must be realistic, and allow the player to control a 'real' character. People working in the field of cultural heritage are keen to stick to reality, as moving away from it would distort the teaching of history.

3D-captured real characters will be controlled through a mobile device in their real environment. There will be no need for additional interpretation of the environment or the costumes to model them, and no more unnatural-looking animations.

When visiting a historic location, visitors can download the associated game on their mobile device. They can then discover virtual artefacts through their device during their visit, and can chose to pick some of them for later use. For instance, while in the armoury, visitors can grab a sword which might prove useful later in the visit, when they are 'attacked' by 'invaders' outside the castle.  Hence, each visitor can control one "real" actor and through a series of missions discover the "real" life in a castle.

In this application scenario, the realism introduced by RE@CT will serve historical facts and give more power to the learning aspect of the proposed game. All the technologies developed during the project will contribute to helping cultural heritage centres encourage people to learn about history.

## 2.2 An interactive application alongside a TV-production

The second application scenario will be developed alongside a TV production. Broadcasters are already showing significant interest in exploiting TV content via interactive media. The BBC for example provides 3D online games alongside popular TV programmes, such as Robin Hood and Dr. Who, as downloads on its website. However, in the current financial climate it is hard for public broadcasters to justify the production costs of expensive high-end video games to achieve realistic-looking virtual characters of well-known actors or presenters.

A second scenario will therefore be developed to demonstrate the use of RE@CT tools for the cost effective production of an interactive game with a known actor or presenter. The genre could be either entertainment, factual or children programmes. The purpose of the interactive application is either to :

   a)  Enrich a TV production (or brand) through an interactive, web-based programme or
   b)  A completely independent production or
   c)  A seamless mix of the previous two

A typical scenario for a) would be to develop an interactive application out of the theme of a TV programme, in the same way as games have previously been produced at the BBC for popular productions as mentioned above.



Figure 1 - Interactive game of the 'Dr. Who' programme

These kinds of applications have the potential to reach audiences that are spending more time online and / or in front of a computer (the "gaming generation"). This option has been explored before in movie and high-end TV productions and is also known as transmedia or cross-platform media, as the programme is primarily designed for the (linear) movie or TV medium and then in a second step ported onto an interactive platform.

Option b) deals with the development of new interactive programme concepts that might be better suited for online platforms, with the same goal as before: to reach new audiences.

Option c) allows for a number of scenarios. The RE@CT technology could be used in a live TV-production set-up. In this case the captured action can be an integral part of a programme. For example, presenters or even well-known figures of public life could be captured in advance of a production and the pre-captured action integrated during the production. For example, a factual programme presenter could be captured whilst talking through a topic and reacting to questions or

actions from a studio audience. Similar techniques could be used to liven up broadcast events such as political debates prior to an election, for instance.

However, because of the nature of the medium, option c) would also allow completely new programme ideas to be captured and streamed live as a TV programme and online as an interactive programme. One example would be in cross-media experiences in which the viewer can watch the programme on TV, but also explore it on a 3D interactive platform at the same time.

# 3   Definition of production requirements

This section describes requirements for each stage of the production methodology, from the video camera system set-up to the outputs provided to the final applications. The production of an interactive application involving the capture of characters using the RE@CT techniques requires a number of these production stages or steps:

a)   Pre-production
b)   Studio capture
c)   Processing and Application production

The purpose of the pre-production is to develop the concept and design of the application, i.e. the story and the ‚set‘ materials etc. Further, the pre-production has to develop detailed plans of what needs to be captured in the capture phase, and how, otherwise not all the required actions arecaptured and in the worst case an iteration, i.e. another studio capture session is necessary, which puts up costs.  In order to avoid this, a detailed plan of the required action needs to be developed and is visualised through a ‚motion graph‘. During the processing and production the motion graph is filled with data and then forms the basic data representation for the animation engine.

## 3.1 Pre-Production

RE@CT 3D characters are animated by concatenating and blending fragments of captured movement together to form novel sequences.  It is therefore essential for the studio capture to ensure: 1) sufficiency in coverage of character movements, anticipating the range of animated movements that will later be generated; 2) that captured movements that will later be concatenated contain suitable "transition points" to enable one movement to flow seamlessly into another. Once captured, the movement fragments may be considered as nodes in a graph, with directed edges in the graph indicating seamless transitions.  This "Motion Graph" is specified manually at a coarse level of detail prior to capture, in order to ensure 1) and 2) are met.

Requirements for full body capture – What needs to be captured?

As a basis, we will focus on the cultural heritage demonstrator. However, the same framework should be set up for any applications requiring re-animated motions.

The following motions have been identified as necessary for the cultural heritage demonstrator, and are grouped by category as shown below.  Motion fragments indicated with * are periodic and will be edited so that they may be seamlessly repeated.  A default casual standing pose will be adopted before and after non-periodic actions, since most of these actions connect either via C1::1 or C1::6 (either directly or via a transitional fragment in C6).

C1: Standing and Looking
1. Standing still (Face: casual)*
2. Standing still and glance to the left
3. Standing still and glance to the right
4. Standing still and looking around in a worried manner (Face: anxious) *
5. Standing still and looking intent/focussed (Face: concentration) *
6. Standing still with regular sword drawn*

C2: Walking and Running
1. Running forward*
2. Running and turning left*
3. Running and turning right*
4. Running and looking left*
5. Running and looking right*
6. Walking forward*
7. Walking and turning left*
8. Walking and turning right*

C3: Jumping
1. Jumping on the spot*
2. Jumping whilst running*

C4: Interaction
1. Climbing onto a wall*
2. Climbing off a wall*
3. Climbing up stairs*
4. Fall to floor from standing and lay still (Face: afraid)
5. Stand up from lying on floor
6. Kneel down from standing
7. Get up from kneeling and stand
8. Open chest whilst kneeling
9. Kneeling ready to open chest

C5: Swordplay
1. Take out sword (from standing)
2. Put away sword (from standing holding sword)
3. Take one step back holding sword (from standing holding sword) (Face: surprise)
4. Fighting stance holding sword raised
5. Fighting stance holding sword and back sword both raised
6. Take out back sword and raise it
7. Put back sword away
8. Swing sword (from fighting stance holding sword)*
9. Swing back sword (from fighting stance holding both sword and back sword)*

To manage complexity, category C5 does not permit certain actions.  For example, it is impossible to be in the fighting stance without having the sword drawn and it is impossible to have the back sword drawn unless in the fighting stance.  It is necessary to draw the sword whilst standing, before entering the fighting stance.  It is not possible to move with the sword/s drawn other than to attack by swinging it/them.

Many of the motion fragments in C1-5 will need to be performed in succession, but are not directly compatible for a seamless transition.  Therefore a further category of transition fragments will be captured.

C6: Transition Fragments
1. Stand-to-walk
2. Walk-to-stand
3. Stand-to-run
4. Run-to-stand
5. Stand-to-jump
6. Jump-to-stand
7. *(Run-to-walk)*
8. *(Walk-to-run)*
9. Walk-to-climb stairs
10. Climb stairs-to-walk
11. Stand holding sword -to- fighting stance holding sword
12. Fighting stance holding regular sword -to- stand holding sword

Transitions run-to-walk and walk-to-run will be captured as a fall-back position but are unlikely to be used as the parametric motion graph will be used to blend between walk/run fragments and synthesise this transition.

The matrix below indicates all possible transitions planned between these motion fragments.
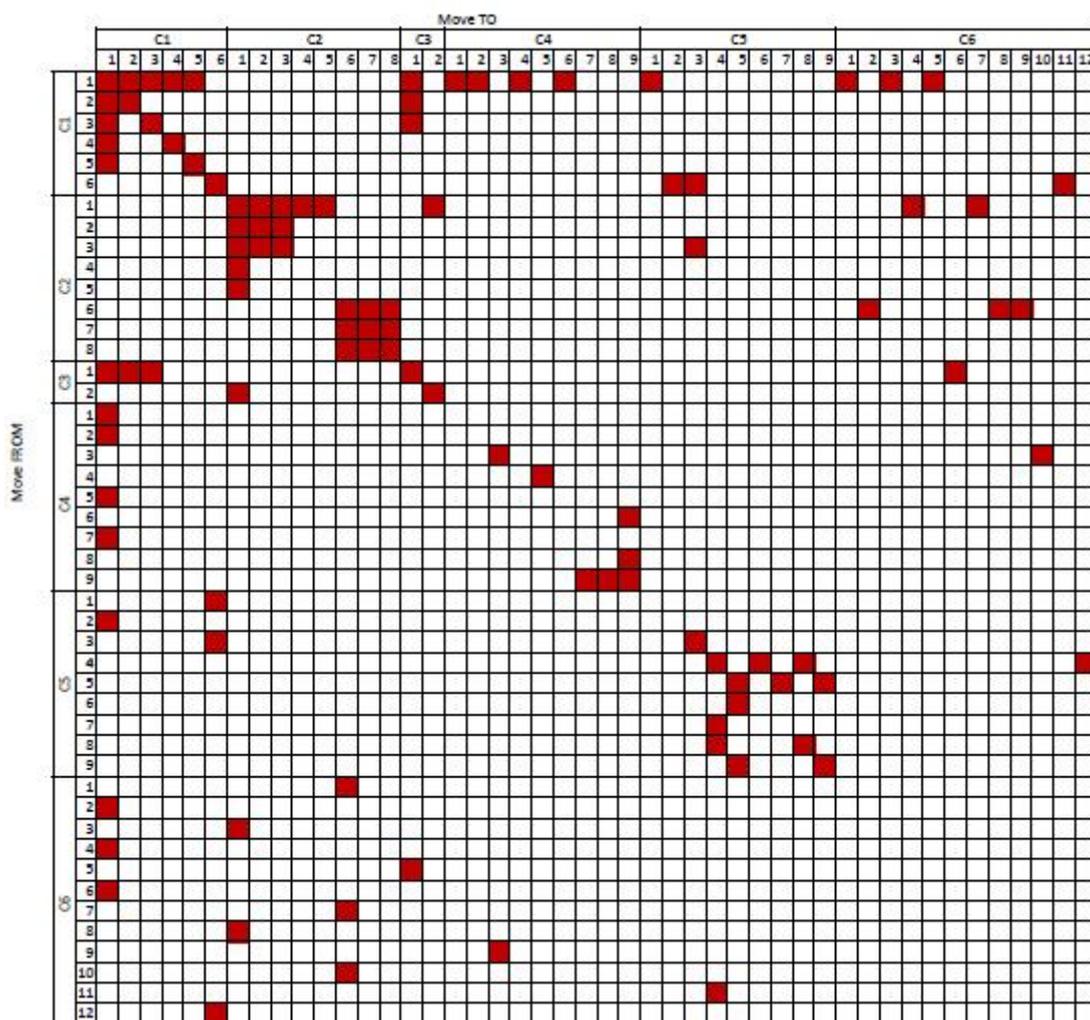


Figure 2: Possible transitions planned between recorded motions

The motion graph below illustrates in schematic form some of the transitions possible between a subset of motion fragments likely to be frequently used in the demonstrator.
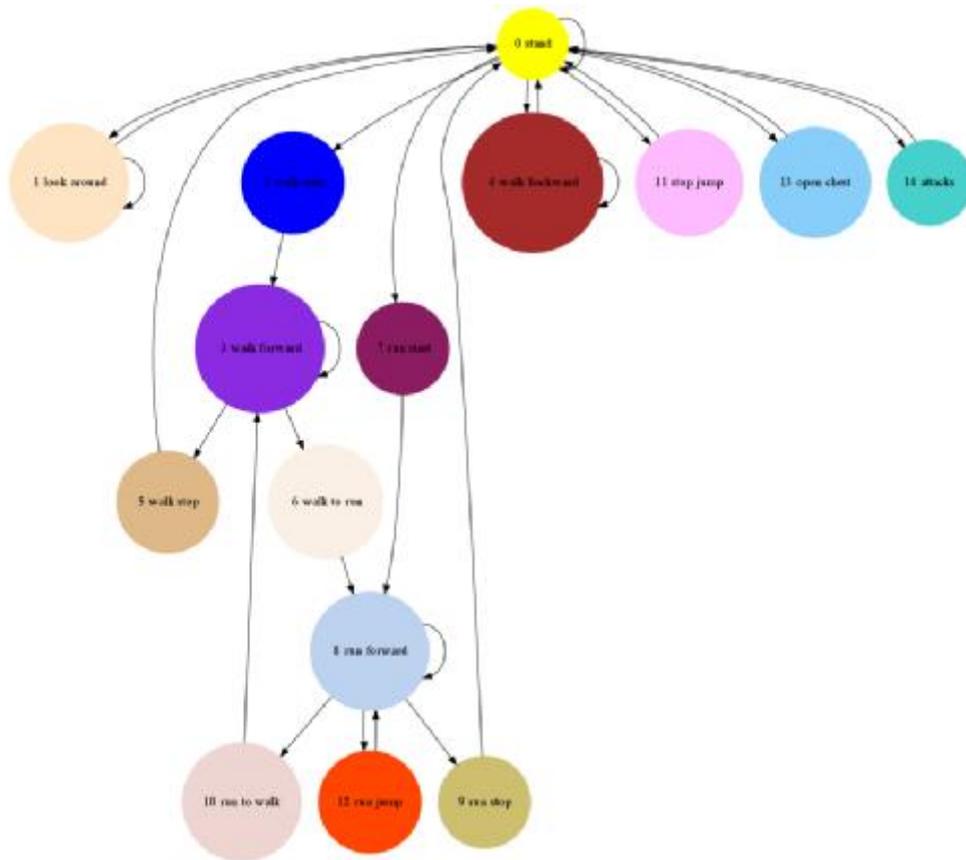


Figure 3: Motion graph in schematic representation from a subset of the possible transitions

**Requirements for full body capture – How will footage be captured?**

A calibrated multi-camera blue screen studio, as described below, will be used to capture each motion fragment in isolation. Four versions of each motion fragment will be captured in costume, for redundancy and purposes of development:

- Regular motion
- Reversed motion
- Regular motion with no props (i.e. no sword, where applicable)
- Regular motion with motion capture markers present

Walls and stairs where used should be covered in blue so as to avoid becoming part of the reconstructed model.

## 3.2 Studio Capture

The studio capture aims to capture all nuances of the action of an actor in the highest possible quality. The capture has a number of sub-systems with specialised tasks:

- The audio capture is based on standard production facilities to record the sound of the action

- The whole body component captures the body language of the actor, i.e. it is framed to capture the entire body
- The head capture is designed to give facial details of the action
- The hybrid motion capture is an optional module that captures body and facial expressions with conventional (marker-based) motion capture technology

### Whole body capture

The capture of the whole body action needs a high visual quality. As specified in D5.1.1 RE@CT uses 10-16 broadcast quality cameras in HD resolution at 25fps as a minimum specification for the capture. The capture system also contains certain processing stages up to a temporally unaligned 3D model of the action. The processing result of each (sub-)module has an effect on the final visual quality of the RE@CT system.

The processing (sub-)modules are:

1. Camera calibration
2. Pre-processing
3. Segmentation
4. 3D reconstruction

The calibration module estimates a set of parameters for the cameras that can be used to relate 3D points to their 2D projection. The quality of that camera calibration has a strong influence on the quality of the final 3D reconstruction: any errors in the camera calibration will lead to problems in the 3D reconstruction. For the visual hull method as described later this is typically articulated by 3D reconstructions becoming eroded and/or missing details like limbs. It is therefore important to get a good camera calibration.

As a rule of thumb the calibration object (or markers) used for camera calibration should be captured at positions that are well distributed over the entire capture volume. That means the calibration object has to be swept evenly through the volume that is used for 3D reconstruction. In order to judge how good the calibration is (or might be) it is important to visualise the quality and the 'coverage' of the calibration during or shortly after the calibration and to allow the calibration process to be repeated if it is not good enough. It is also good practice to do a calibration run before and after a capture. That allows the later 'back-up' calibration to be used if some camera parameters changed during the capture, e.g. if a camera was accidentally knocked.

The pre-processing module performs a number of image processing steps, that partially depend on the camera and capture technologies used. Broadcast quality cameras will usually deliver image data via HD-SDI in YCrCb colour space which needs to be converted to RGB colour space. This module will also correct camera artefacts such as lens distortion.

The segmentation module classifies the pixels of a digital image into foreground and background, using a process known as segmentation or matting (in the film industry) and keying (in broadcast). The result of this process is stored as the 'alpha-channel' of an image. The main application for keys or mattes is in compositing, for example to exchange the background of a scene with a different image. In our application, the results of the segmentation process are used as an input to the 3D reconstruction module.

The 3D reconstruction module computes a 3D model of each frame of the action. A quick algorithm to compute a 3D model is a visual hull computation. This algorithm uses only the camera calibration data and the scene segmentation. The resulting 3D models can be used for pre-views and/or on-set visualisation as it can be computed in real-time. The visual hull model can also be used as a starting point for more elaborate methods.

**Head capture**

For high-quality face close-ups such as dialogue scenes there is a dedicated head capture pipeline in RE@CT. It comprises two parts:

1. A pre-modelling phase where detailed 3D models of the actors are acquired using passive stereo techniques developed at HHI. The capture hardware comprises off-the-shelf SLR cameras, lighting (i.e. flashes and diffusers) and tripods for mounting these. The equipment is sufficiently compact to be shipped to a filming location for on-site capture. It can be set up by two people within an hour.
2. A studio phase where the facial performance of the actors is captured with dedicated head cameras which pan, tilt and zoom capability in order to keep a close-up view of the actors' faces from multiple viewing directions. Dedicated head capture can happen simultaneously with full-body capture or separately.

The detailed 3D models are then used to analyse the online footage and render new content based on the studio capture footage. Technical details of these processes are specified in deliverable D5.1.1.

The content to be captured in the studio phase determines what has to be captured in the pre-modelling phase as, for analysis, the 3D models are matched to the studio footage. The more similarities there are, the more robust is the analysis result. HHI is currently developing a standard catalogue of facial expressions for the pre-modelling phase which aims at capturing a sufficient amount of variability to analyse most types of footage captured in the studio phase. For specific scenarios, e.g. a fighting sequence, it is advisable to include related facial expressions in the model database. The size and variation of the model database required for analysis is a matter of on-going research.

## 3.3 Pre-visualisation

During the production phase in the studio it is important to get a quick feedback that the captured content meets the quality standard both from a technical and creative point of view. For this purpose it is essential to get

a) a play-back of the captured material
b) a pre-view of the captured 3D action
c) a visualisation of any virtual object either dynamic or static

The play-back is possible through our multi-camera capture system. For a pre-view, a fast 3D-reconstruction is computed using a visual hull technique, as described above.

For an on-set visualisation of virtual objects, we are planning to use an on-set projection system, as described in D5.1.1. This system produces a view-dependent projection for the actor's point of view. The actor can interact with this projection, e.g. by looking at a virtual object or even 'touch' it.

# 4   Specification of software layout and design

This section presents the different steps of the processing performed to obtain the re-animated mesh from the perspective of applied methodology. We present the capture system, explain a practical methodology for preparing the meshes to be re-animated and finally present some practical hints for the reanimation scheme.

## 4.1 Capture system

The capture system allows the capture of action from one or more actors. The capture components are mainly built from standard production components. For the capture of sound, standard close-up radio microphones and audio recording equipment are  used. Components specifically configured or developed for the RE@CT project include the multi-camera sub-systems for a) whole body (visual) capture, b) facial detail capture and c) hybrid motion capture.

The capture system is operated through a GUI, as depicted in Figure 4. The GUI allows all the cameras in the system to be monitored in a single overview display. Also, live colour histograms can be displayed to help the camera set-up. A capture can be started by pressing the 'record' button. This starts recording 2s prior to the event (a feature which is important for capture of 'live' events). During the recording the GUI displays information about the status of the recording disks and informs the operator of any bandwidth problems that could potentially lead to frame drops.



Figure 4 Capture GUI during first RE@CT test capture

### Synchronisation of cameras

The visual capture components for the whole body, facial details and motion capture are all based on video imagery. The RE@CT processing pipeline assumes that the images are temporally co-timed. That means the images from multiple cameras should be synchronised and the shutters should all be open at the same time. Professional broadcast cameras provide a gen-lock facility allowing synchronisation of their capture frequency to a 'master' signal.

However, gen-lock only guarantees that the cameras are running with the same capture frequency. In particular cameras of different types can have very different timings or phases, although most broadcast cameras allow adjustment of the timing phase.

Figure 5 shows the GUI for a tool that has been developed to check and set the camera timing. The LEDs are flashing with a frequency higher than the gen-lock frequency and are synchronised with the gen-lock - this is done through a PC having exact timings of the incoming video packages. The relative

phase of the blinking LEDs and the gen-lock is now altered until all LEDs appear 'lit'. For another camera, the camera phasing is then adjusted until all LEDs appear lit.
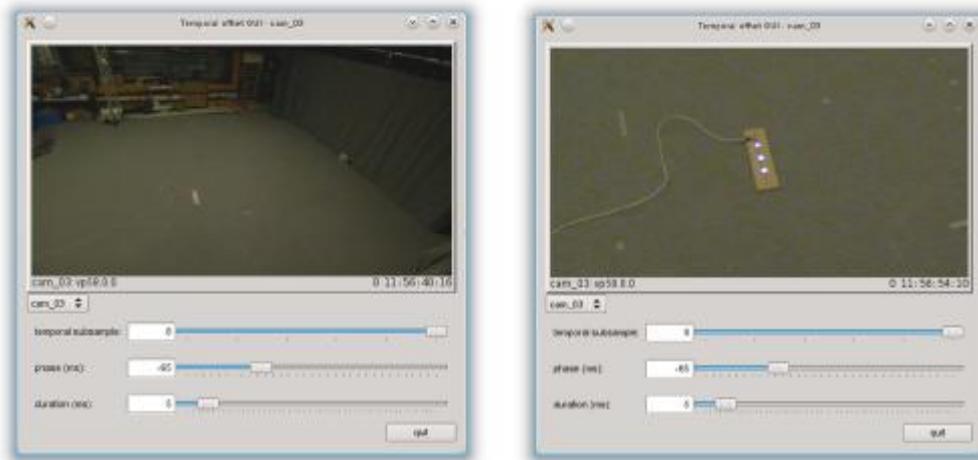


Figure 5 - LED board and GUI for camera synchronisation

## Whole body capture

For the calibration of the whole body capture system, a camera calibration technique was developed which can be used to calibrate both the widely-framed whole body cameras and the more tightly zoomed-in facial detail cameras.

The calibration object, shown in Figure 6, was designed using colour-coded LEDs. Both ends have a differently coloured LED (red or green, the rest are white). This allows the identification of LEDs even when the object is only partially visible in the image. This feature is particularly useful when actors are captured with cameras of different framing, i.e. cameras framing the whole body and facial detail.



Figure 6 A colour-coded LED calibration object

Once the marker positions have been identified a bundle-adjustment process is used to compute the camera parameters. A GUI has been designed to supervise this step and adjust the parameters for the calibration, as depicted in Figure 7.

Figure 7 GUI for camera calibration

## Facial detail capture

Facial detail capture with pan-tilt-zoom (PTZ) cameras can happen simultaneously with full body capture or in a separate session. The PTZ cameras are pre-calibrated and provide accurate data about the pan/tilt and zoom parameters, as these parameters are used to control the motorised PT-head and the zoom lens. Alternatively a manually-operated camera can be used together with a system previously developed by the BBC which is based on a small secondary camera which is fixed orthogonally on the primary one. The secondary camera points to the studio ceiling where retro-reflective, uniquely identifiable markers are installed. These are used for extrinsic calibration. For intrinsic calibration, the zoom lenses of the PTZ cameras are equipped with an electronic feedback system. An online calibration refinement, which operates on the captured footage, may be necessary and is being developed by HHI.

HHI's facial tracking system uses the high-resolution models generated in the pre-modelling phase to track the head in the PTZ camera footage and to add consistent depth information for re-rendering. This is achieved by registering images from the pre-modelling phase (which have depth information) to the video footage. Matching is performed in two stages: the first one is rigid and performs pose estimation. The second stage is non-rigid and helps to match facial expressions which are not represented in the database. This process is illustrated in Figure 8. Finally, temporal consistency between frames is established. The matching process is automatic. If corrective user-interaction proves to be necessary the required tools and user interfaces will be implemented.

Figure 8 Matching of a database image, capture at the pre-modelling stage, to a video frame from a PTZ camera

### Hybrid motion capture

The hybrid motion capture system will use pre-existing tools from the commercial motion capture software products from OMG as much as possible. These products provide user interfaces to control and set-up the camera system, adjust the settings of each camera, perform camera calibration, record video and inverse kinematics, and present real-time feedback to the user for data quality assessments. A brief overview of the user interfaces follows.

Once the hardware is in place, in order to maximise the quality of the motion capture data, it is necessary to optimise the camera parameters depending on the environmental conditions. For example different illuminations might require a different setup. Figure 9 shows the interface presented to the user during system setup. The left panel enables the selection of one or multiple

cameras. A live preview of the videos and 2D marker tracking results is shown by the central panel where it is possible to select and zoom into specific parts of an image to assess the quality of the video better. On the bottom left another dialog box contains tools to adjust the camera parameters such as signal gain, intensity of infra-red strobe lighting, and marker detection and tracking thresholds. Based on an initial quality assessment it might be desirable to re-adjust the camera positions. The goal is to improve the camera view overlap, as at least two cameras must see each point in the volume. Also it is desirable for two infrared cameras not to face each other as infrared strobes in the field of view saturate the pixels in that image area.
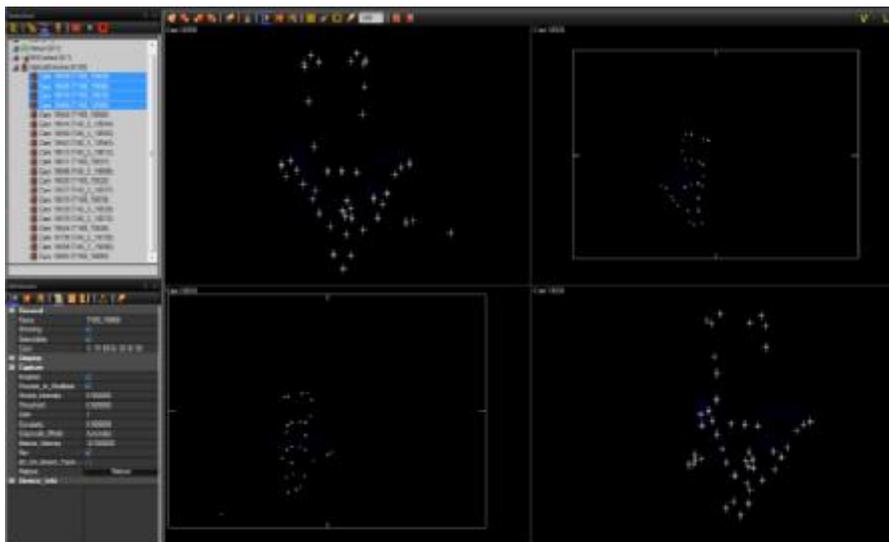


Figure 9: User interface for camera setup. Left, two dialog boxes allow selection of one or more video cameras (top-left) and parameter editing (bottom-left). The rest of the UI displays data previews where the bright pixels represent the marker positions.

When the user is confident that the setup is appropriate to the studio environment, they can move on to the camera masking and calibration steps. Camera masking is a thresholding operation performed on an empty volume to disregard areas of the image plane occupied by shiny objects such as the IR strobes of the other cameras. Camera calibration determines the relative position of each camera, estimates the internal camera parameters as well as the parameters of the lens distortion model. To calibrate the cameras, the user has to switch the application into calibration mode and use a calibration device (an LED wand) supplied with the system. The wand is waved in the capture volume; the application detects the position of the LEDs in multiple frames and uses the data for calibration. Figure 10 shows the user interface in calibration mode. As the user waves the wand in front of the cameras, smears start to appear in the preview panels representing valid wand detections, i.e., detections that will be used to compute the calibration results. The panel on the bottom right also shows a count of the valid wand detections.
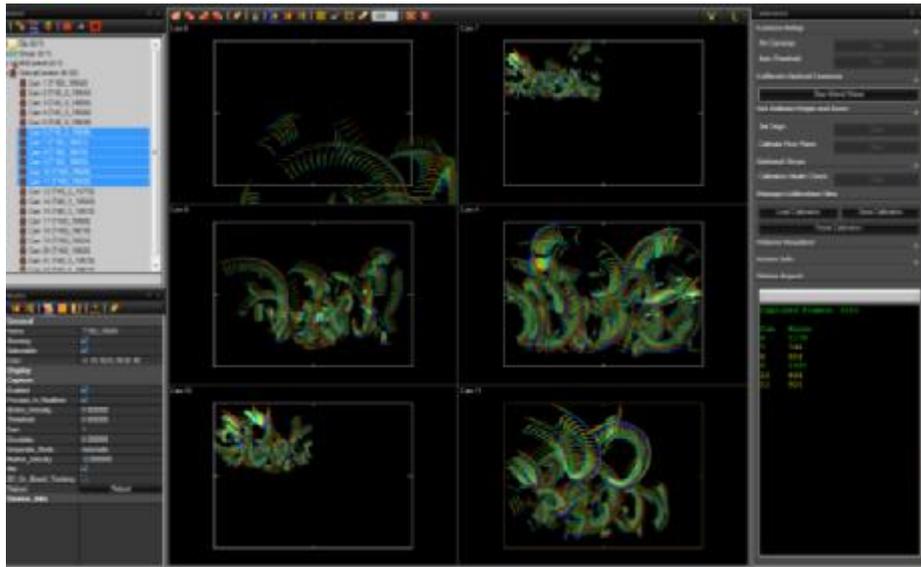
Figure 10: User interface for camera calibration. Right dialog: camera calibration controls and system status. Central panel: calibration data preview. For each camera the valid wand detection results are presented as multi-coloured smears on the screen. This serves as user feedback regarding the calibration data coverage.

The final step of the setup procedure is to create kinematic models of the actors. To this end a Range of Motion (RoM) trial is recorded, where the actor performs a set of predefined moves. The raw data is loaded in the application and the marker positions are reconstructed. Then the user switches the application to "actor setup" mode (see

Figure 11) and is presented with a set of dialogs and editors with tools to define, modify, and manage models for multiple actors. The application comes with preloaded template models and with algorithms to calibrate the models automatically, based on the RoM data.
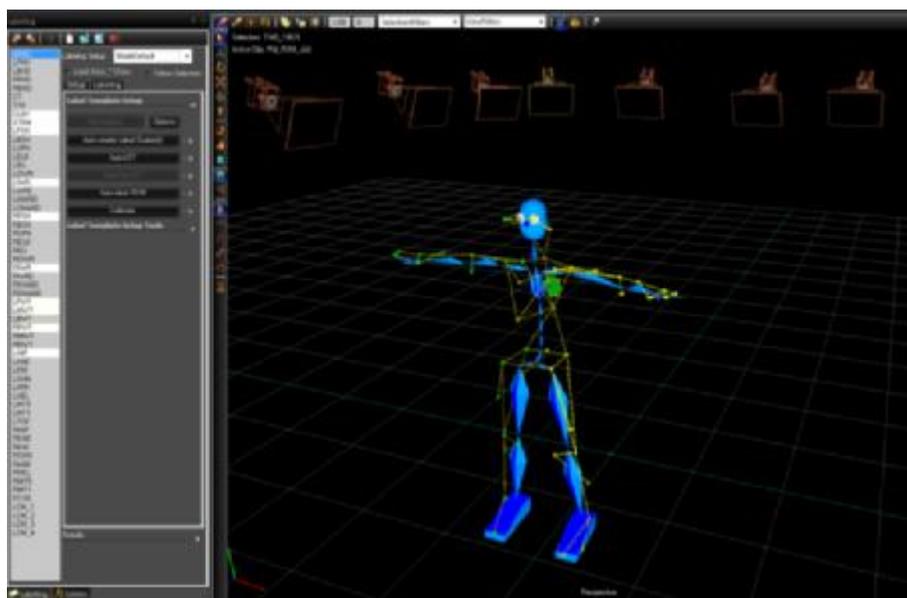


Figure 11: User interface for actor model setup. Left dialog: user controls to manage subject models for multiple actors and to invoke subject calibration and marker labelling algorithms. Central panel: actor model preview. The toolbars enable different manipulators to change the relative pose of "bones".

For future use, camera and actor setup can be exported and imported as a set of XML files. Given these models, the user switches the application to "Capture mode" and is presented with a 3D workspace (Figure 12, left) showing a real-time preview of the actor motions and surface data. A dialog box (Figure 12, right) includes controls to start and stop capture, specify file names and capture paths, manage the active characters in the scene, preview the last take and control the state of the synchronisation with external gen-lock and time-code signals.
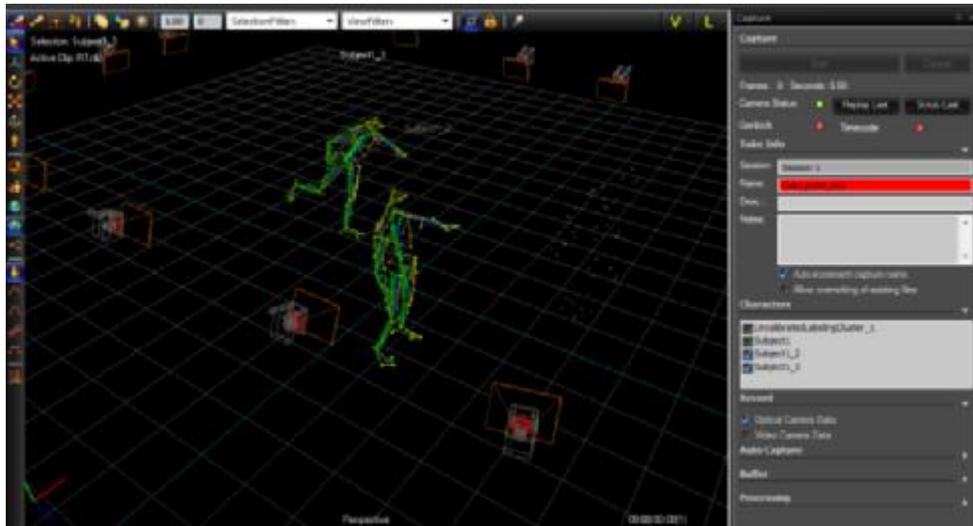


Figure 12: Data capture user interface. Left: 3D workspace for real-time data previewing. Right: dialog box with capture controls.

## 4.2 Tools for Performance analysis

Once sequences are captured, several processing steps need to take place, with the goal of providing a single 3D model of the person. There are four identified analysis steps to build a complete 3D model template: correspondences over the sequence, semantic annotation, detailed appearance analysis and retrieval, and facial expression acquisition. These analysis steps will happen offline as a data preparation step for building a motion graph, and will each be tackled by a corresponding software component.

The components corresponding to each analysis algorithm share some common traits, in that they take the data previously captured as input, in the formats specified in D5.1.1, and can be viewed as a semi-automated process requiring a GUI visualization of 3D data and algorithm results. The GUI associated with each component will typically look like the interface shown in Figure 14, with a central 3D visualization area, showing the 3D input information used by a module, outputs generated by the associated analysis algorithm and algorithm execution, specific 3D data and visual annotations to diagnose the algorithm performance. Each will include a sequence chooser to pick a sequence in the acquisition database, a time line navigation slider enabling to select data associated to a given time stamp, and a specific pane containing the parameters and execution control buttons of the analysis algorithms. Another pane will typically be included to switch and control different visualization parameters, such as turning on/ off a specific input /output data layer.

Depending on the findings of the project, some components may interact and use the results of others to achieve their goal. As such some components may be built in the same application, in which case their corresponding parameter panes will be tabbed in the GUI for easy access to their

functionality. A master pane will control the interaction between components. General menus will enable the general sequence database to be loaded and parameters to be set which cover application-wide functionality. We describe the specifics of each component in more detail below.
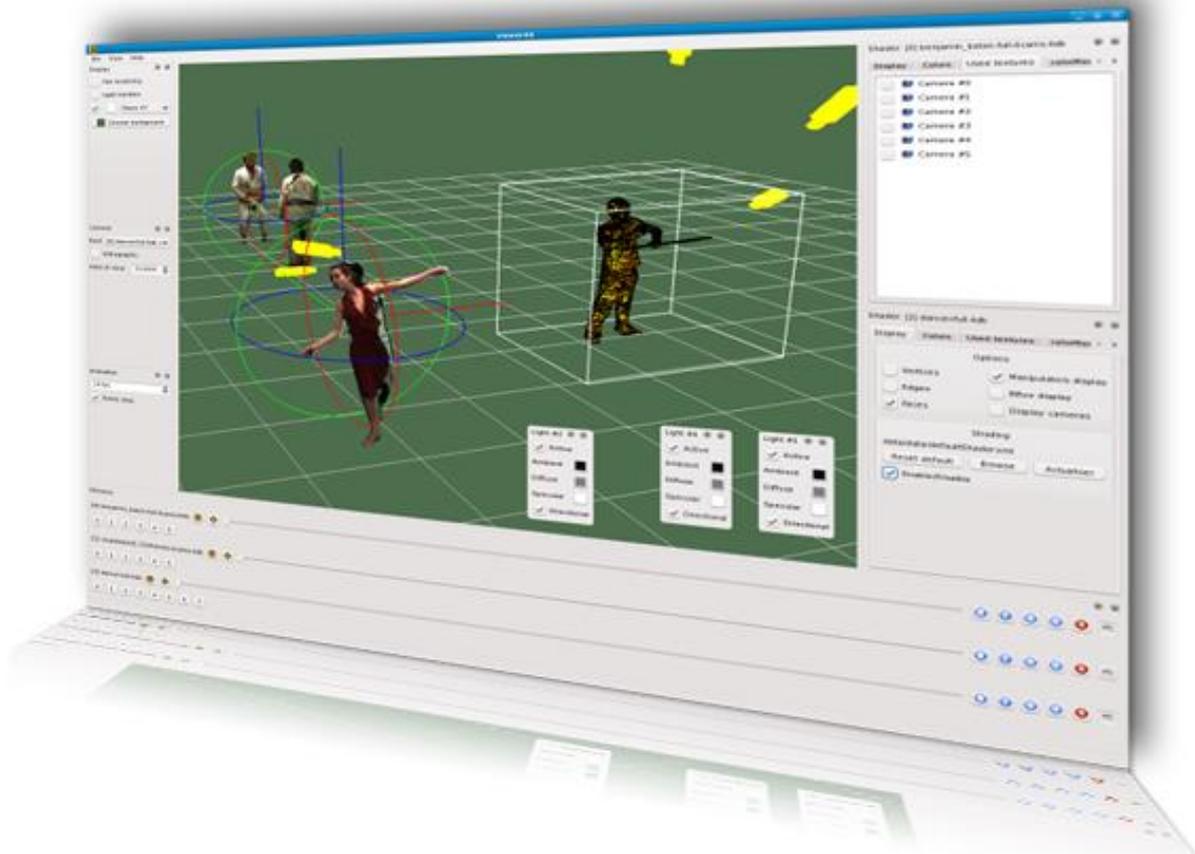


Figure 13 - An anticipation of the Performance Analysis Software Interface

### Establishing Surface Correspondences across the Sequence

The task of establishing correspondences in the sequence will use a surface tracking/matching algorithm. It will predominantly use the full body capture data as input but may also use other data sources among those acquired, depending on the exact details of the final technique. Such algorithms usually proceed in two distinct steps. The first step consists of choosing a 3D template model whose shape and quality are representative of the observed surface deformation in the sequence. The second is to fit it to data from the sequence.

The first step will typically be a semi-automated analysis  to be performed interactively by an expert who will identify, among the sequences captured, the frames that can provide the best 3D model  to represent the body shape and be used as a template initialization. Once such frames are identified, an initial 3D reconstruction is retrieved from the captured data for this frame. A set of tracking/matching parameters should be set in accordance to the geometry and topology of the tracked model in its dedicated pane, where the correspondence-finding process will be launched using a button. Since the retrieval and tracking of such a model may require several attempts, the GUI developed for this purpose will allow the 3D models obtained to be visualized, built and compared. The output will consist of an animation model with fixed topology, with coordinates deformed temporally to optimally represent the corresponding observed 3D shape.

### Semantic Annotation and Rigid Part Identification

A semantic analysis software component will be dedicated to automatic semantic annotation and sequence segmentation, as an aid to label and segment the sequence into relevant parts for 3D modelling and motion graphs. Algorithms will be designed to extract this information primarily from the tracked/matched sequence previously analysed, but may also use the 3D capture data, or use images directly, depending on project findings. Similarly to other components, the semantic analysis will have a dedicated pane with the parameters controlling its execution and a launch button. In the GUI visualization, the annotations, such as rigid part body segmentation, events or actions, will be made visible typically by colour coding the output mesh, the capture timeline, and/or the set of acquired sequences, which will be toggled on or off in the component pane using checkboxes.

### Appearance Analysis

The main goal of appearance analysis is to provide the 3D model with a refined, precise appearance model inferred from the data, to allow very accurate sequence playback and characterization. This component will typically use the results of surface correspondences, captured data and images as input. It may also simultaneously perform geometry refinements to achieve better estimation of the appearance. The result of this step will typically be visualized in the 3D pane of the GUI, taking the form of improvements to rendering such as meshes texture mapped with the estimated appearance.

### Facial Expression Synthesis

The detailed head/face footage can be used to re-render new facial close-up sequences. For the new footage, the viewpoint can be changed freely and continuously between the positions of the dedicated head capture cameras and, to a certain extent, beyond their position. The footage can also be re-combined temporally in order to obtain new sequences. Here, we adopt the motion-graph framework being developed by the University of Surrey. To guarantee smooth transitions between the motion segments, the exact transition points have to be optimized and transition motions have to be computed automatically.

## 4.3 Preparation and processing of motion graphs

The preparation of the full body motion graph is dependent on the completion of the 3D reconstruction and temporal alignment of the mesh.  At this stage the mesh sequence for each motion fragment will be edited so that periodic motion fragments contain several full cycles of action, and non-periodic motion starts and stops in the reference pose.

Once the mesh sequences have been edited, a colour-shape histogram or similar 3D shape and appearance descriptor will be computed for each frame. Fragments that may be concatenated according to the motion graph specified in pre-production (e.g. C1::1 to C6::1) will then be processed to determine inter-frame similarity across both clips. A distance matrix will be computed between all pairings of frames across both sequences, using the extracted descriptors. The matrix shown in the figure below was computed from a walking sequence, and has been thresholded to indicate potential points for sequence concatenation.
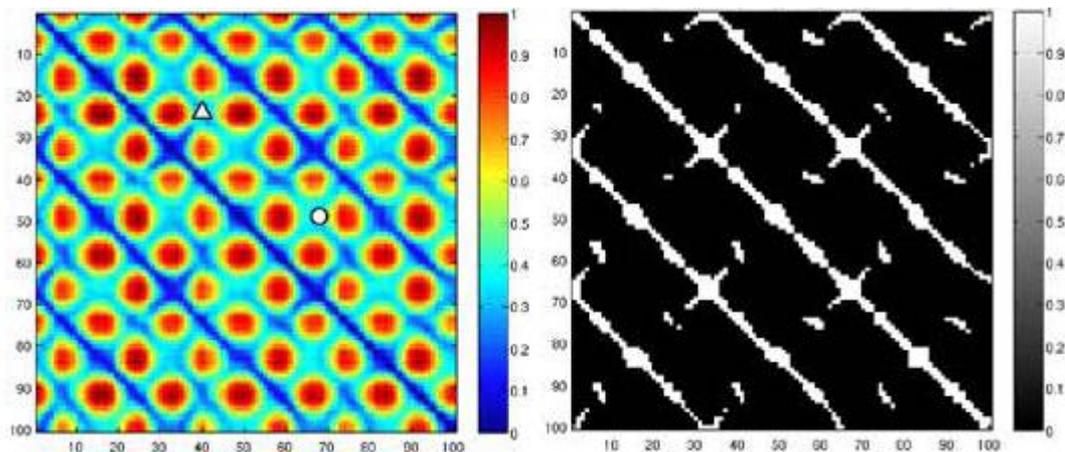
Figure 14 - Example of a distance matrix for a walking sequence (left: original, right: thresholded)

These transition points will then be output in the motion graph XML specification outlined in detail within RE@CT deliverable D5.1.  Due to the real-time nature of the parametric motion graphs, further computation of this data must be performed during the animation synthesis itself. This discussion is beyond the scope of this document.

## 4.4 Application development

This section presents the first design for the prototypes and demonstrators to be built by the project to support the novelties introduced by RE@CT.

### The interactive application for use in a public space

Once the data exchange format has been refined, the interactive application will be developed based on the technical specification described in D5.1.1.

This application will be run on mobile platforms.  For the first prototype , everything will be developed on an iPad ($3^{rd}$ generation). The game engine Unity3D allows deployment on such platforms and eases the process of porting to other platforms such as Android or even game consoles. The general development process will be phased in 6 steps as follows :

- Design of the application: define the functionalities, the tree diagram, the different interfaces
- Develop each of the designed blocks of the application (Network communication, Reanimation tool, Rendering process, etc.)
- Perform unit tests for each developed block.
- Gather all units into  a single intermediate prototype and perform tests
- Integrate graphical interfaces, etc.
- Release the final prototype

Both Unity 3D and iOS native APIs will be used to develop the first prototype, whilst also taking into consideration future compatibility with other platforms such as Android.

Several versions of the prototype will be distributed internally to the consortium to allow for testing and further development, and optimization of other components of the project.

### The interactive application for use alongside a TV-production

This demonstrator will use the same development frameworks as mentioned before, targeting PCs as widely available user platform. The interactive user-facing application will be developed with a RE@CT platform (Unity3D or similar).

For a demonstrator used during a TV production rendering tools developed within RE@CT will be used.

## 4.5 User interface

The first scenario (serious gaming) is aimed at the general public, so this demonstrator will focus on simplicity. The images below illustrate user interface examples for the outdoor augmented reality game. Later, during the development process, we will provide other previews of such interfaces.



Figure 15: Loading screen example

Figure 16: Main menu example



Figure 17: Choose your weapons

Figure 18: Historical pitch: Description of the missions



Figure 19: Outdoor Augmented reality view

Figure 20: "Real" character motion control

# 5   Identification of non-project-related components to be used

The RE@CT project will use a set of external tools already available off-the-shelf or from the project consortium partners, in addition to those being developed within the project. In this section we provide a list of those existing components and briefly describe their proposed use within the project.

## Tools for mesh simplification

3D reconstruction will be performed at a high-level detailed resolution by the tools being developed within the project, as precision and accuracy are essential to avoid the loss of data. However, real-time authoring tools and rendering engines require 3D data sets to be optimized and require the number of polygons to be as low as possible. Moreover, all should be optimized for the targeted platforms, therefore meshes will be post-processed.

At different stages of the process, mesh simplification can be used to allow for the usability of the meshes on mobile platforms. The Wrapit tool will be used internally to 3DSMax.

## Modelling tools

In addition to motion control, which will be one of the key activities of the project in WP4, developing a demonstrator will require the setting up of a complete environment and the provision of 3D models, separate from controlled characters. Those models with be built using 3DSMax. Depending on the chosen 3D environment, we propose to use architecture-related software from AutoCAD to allow for architectural cultural heritage modelling. Ground plans and photos might be used to recreate realistic environments.

Rendering engines

The RE@CT project aims to provide two different types of demonstrators, one related to mobile platforms and the other to live TV production. Rendering tools will be adapted to the targeted application and platform.

In the case of mobile platforms, we plan to use the Unity3D game engine. This engine is a cross-platform development kit, available for Windows, Mac OS X, Linux (availability announced but not yet launched), Android, and iOS. We plan to build a demonstrator around the Android and iOS mobile platforms.

The use of these platforms will enable the project to follow state-of-the-art rendering technologies throughout the project. Unity3D is ready to be released in its fourth version. It brings major novelties with respect to rendering such as real-time dynamic shadows. Major improvements will be also be made on fluidity and workload on mobile GPU, which should allow for higher resolution meshes to be run on those platforms.


Augmented Reality technologies

Given the proposed scenario, we plan to use augmented reality technologies, which fall into two categories as follows:

- A marker-based technology for the board game. Several commercially available solutions are being considered, and a final choice will be made once the demonstrator design has been finalised.
- A geo-localized approach, based on the use of the inertial sensors and GPS. We plan to use tools developed and owned by Artefacto, which make use of embedded sensors on mobile platforms.


# 6  Conclusions


The aim of this document was to specify the use case scenarios for the technology developed by RE@CT. The project will focus on two specific scenarios, one in cultural heritage and one in broadcast. Both will be developed into demonstrators.

Further, the document lists the production requirements and the interfaces for the professional production crew.

This document will be used within the lifetime of the project as a guideline for the development of the functionality of software modules and the project demonstrators. Details and changes in the structure of the technical project components will be covered by deliverable D5.1.2 'Update of specification of platform, animation modules and interfaces', due in month 20 of the project.