D8.4 Report on Experimental Production Scenario Results

Grant Agreement nr	780470
Project acronym	SAUCE
Project start date (duration)	January 1st 2018 (36 months)
Document due:	31.12.2020
Actual delivery date	31.12.2020
Leader	FA
Reply to	jonas.trottnow@filmakademie.de
Document status	Submission Version

Project funded by H2020 from the European Commission

Project ref. no.	780470
Project acronym	SAUCE
Project full title	Smart Assets for re-Use in Creative Environments
Document name	D8.4 - Report on Experimental Production Scenario Results
Security (distribution level)	PU
Contractual date of delivery	M36
Actual date of delivery	31.12.2020
Deliverable name	Report on Experimental Production Scenario Results
Type	R
Status & version	Submission Version
Number of pages	33
WP / Task responsible	FA
Other contributors	TCD, DNEG, UPF
Author(s)	Volker Helzle, Simon Spielmann, Jonas Trottnow
EC Project Officer	Ms. Adelina Cornelia-Dinu, Adelina-Cornelia.DINU.@ec.europa.eu
Abstract	This deliverable documents the experimental productions carried out within the SAUCE project. For each production it explains which technical novelties have been tested, which creative possibilities have been enabled by them and which coordination details came along with them. Each experimental production documentation will also result in a best-practice evaluation.
Keywords	Experimental Production, Use Cases, Evaluation, Virtual Production
Sent to peer reviewer	Yes
Peer review completed	Yes
Circulated to partners	No
Read by partners	No
Mgt. Board approval	No

Document History

Version and date	Reason for Change
1.0 23-11-20	Document created by Jonas Trottnow
1.1 18-12-20	Version for internal review
1.2 22-12-20	Revisions in response to review: final versions submitted to Commission

Table of Contents

1 EXECUTIVE SUMMARY

Within SAUCE, multiple experimental productions have been executed, testing the SAUCE developments in real production environments.

All experimental productions have been documented extensively, not only in text form but also in behind-the-scenes and summary videos. Multiple of these videos have been produced by Filmakademie and published publicly. A short-film, a music video and an interactive visualization for a holographic display have been produced with SAUCE technology.

Due to the Covid-19 pandemie the original experimental production plans reported on in the 'Interim Experimental Production Evaluation Report' (D8.2) had to be modified. Most of the staff has moved to HomeOffice for a majority of 2020, making collaborative productions almost impossible. Studio facilities were closed. Student and commercial productions have been cancelled or delayed. In addition, productions that were still possible in 2020 had to be shrunk down to a minimum amount of people on site and a minimum amount of time spent together with others. This often prevented to add SAUCE evaluations to running productions.

Nevertheless new approaches for the experimental productions have been found. Dedicated SAUCE experimental productions have been developed. Material and concepts from e.g. student productions has been transferred to controllable, separated experimental productions for SAUCE. In addition LED wall technology was added to many of the experimental productions, introducing new possibilities both for reacting to the new production requirements within the pandemic and evaluating SAUCE technology in the context of this trendsetting and much discussed new way of working in a virtual production environment.

2 BACKGROUND

All aspects of SAUCE developed in WP3 to WP7 shall be tested and demonstrated in experimental productions. This is one aspect of the evaluation carried out in WP8. As reported in D8.2, three main topics have been identified to be addressed by these experimental productions:

1. Light fields are smart assets already in themselves, directly when being captured. They are versatile usable and reusable. Within the SAUCE project (WP3), the entire pipeline for using and creating light fields is addressed. Starting with the development of camera and capturing technology, through preprocessing to post-production tools.
2. The search and transformation frameworks developed by Double Negative and The Foundry (WP4,WP5 & WP7) make it possible to reuse and search through existing asset databases without the need of manually tagging all of them. This transforms older, probably unused assets to valuable smart assets, potentially saving loads of time in the initial stages of media productions, where initial layouts and fundamental decisions are addressed.
3. In order to make characters and crowds smarter, technology is developed to automatically animate a rigged character based on machine learning networks that learned how e.g. humans move (WP6). Characters are made aware of their surrounding (scene aware) and are thereby enabled to autonomously navigate in the environment. Additionally rigs and animations are made reusable by automatizing retargeting between different characters and stylization of animations. On top of that crowd tools are developed to author crowd animations more easily and intuitively.

This deliverable is a follow up on the 'Interim Experimental Production Evaluation Report' (D8.2). D8.2 already summarized the experimental productions having been carried out in the first half of the project. This deliverable adds the productions that followed thereon and puts them in perspective with future usage scenarios.

3 INTRODUCTION

This deliverable documents the experimental productions carried out within the SAUCE project. For each production it explains which technical novelties have been tested, which creative possibilities have been enabled by them and which coordination details came along with them. Each experimental production documentation will also result in a best-practice evaluation.

3.1 Main objectives and goals

Main objective of the experimental productions was to evaluate developments from the SAUCE project within real-world production environments. The capabilities and possibilities have been put to the test with the goal of developing new pipelines and workflows for utilising SAUCE technology in production and coming up with best-practise suggestions. The intention is also to evaluate the potential of the included technologies for future projects and productions.

3.2 Methodology

By going hands-on with the developed aspects of SAUCE within real production environments the true potential and challenges become apparent. Combining the novelties of SAUCE with existing and established workflows and technologies is the key for this.
The experimental productions were carried out under realistic, professional working conditions in established media pipelines.
Wherever possible, external contributors coming from industry were invited to participate in the experimental productions. They were trying to apply their knowledge and experience from their previous work to the experimental productions using SAUCE technology. This ensured that an unbiased opinion and workflow was generated. As an example the "Unfolding" shoot was directed by the well-known director of photography (DoP) Matthias Bolliger. The light field elements shoot was accompanied by the Dneg VFX supervisor Simon Pabst.

3.3 Terminology

FA - Filmakademie Baden-Württemberg
FO - The Foundry
USAAR - Saarland University
Dneg - Double Negative
UPF - Universitat Pompeu Fabra
ML - Machine Learning
LF - light field

3.4 Self Assessment

According to the self assessment plan (D2.1), WP8 shall "test the prototype technologies and toolkits in a series of experimental productions and evaluate their performance in realistic contexts of professional use". This has been achieved by realizing multiple experimental productions.
The initially identified student projects were not able to evaluate all aspects of SAUCE during production as these had to be scaled down due to Covid-19. Nevertheless the material and scenarios from these productions have been transferred to dedicated and controllable experimental productions. Remaining aspects have been addressed by experimental productions carried out by the SAUCE partners UPF and TCD.
WP8T2 is successfully reported on herewith, only a small amount of SAUCE outcomes could not be included into experimental productions due to the restrictions of the Covid-19 pandemic.
Creatives at FA as well as professional collaborators outside and inside SAUCE have been included into already executed experimental productions to be able to judge the usefulness and usability of the tools from a professional point of view.

4 Overview & COVID-19 Impact

4.1 Overview of experimental productions

Following is the list of experimental productions that have taken place. Due to the Covid-19 pandemic the original plan reported on in D8.2 had to be adjusted significantly. New experimental productions have been established to cope for the lag of some of the planned ones. Following is the adjusted table of experimental productions mapped to the included workpackages and tasks.

Experimental production abbreviation mapping:
EL: FA Light field elements (reported on in D8.2)
UF: FA Unfolding & Unfolding 2.0
LCM: LED Cave Production Mannheim with FA 50 Megatons Assets
LWF: LED Wall Production at Filmakademie Studio with FA Woody Assets
DIE: DNEG Search & Transformation Framework Internal Experiments incl. Foundry Asset Reuse*
VG: UPF Videogame
DC: TCD DublinCrowds
DNC: DNEG Crowds

* The internal experiments by DNEG and Foundry are being reported on in D8.5

Topic	Partner	Experimental Production
WP3		
WP3T1 Smart Assets via Lightfield Capture		
- LF rig general (live preview, data handling, rigging, ...)	USAAR	UF & EL
- LF 5D aspects	USAAR	UF
- LF Pipeline		
- - Calibration	BUT	UF & EL
- - Debayering	USAAR	UF & EL
- - Colour chart based Colour Correction	TCD	UF & EL
- - Intrinsic Colour Correction	UPF	UF & EL
- - Rectification	BUT/USAAR	UF & EL
WP3T2 Transcoding of Lightfield Assets		
- Preprocessing (Denoise, Super-Res, Hot Pixel, Antialiasing)	TCD	UF & EL
- Depth Estimation	TCD	UF
- Tilt-shift rendering	TCD	UF
WP3T3 Acceleration of Lightfield Processing		
- SLAM++ based calibration & rectification	BUT/USAAR	UF & EL
WP4		
WP4T1 Smart Search Framework Development		
- Framework in general (UI, supported file formats, modularity, chain classifiers)	DNEG	LCM & DIE

- Smart Search Framework & FLIX Integration	DNEG	DIE
WP4T2 Asset Validation Tools		
- Deprecated asset types (description/discovery)	DNEG	DIE
- Transformation / upgrade integration	DNEG	DIE
WP4T3 Semantic Labelling Tools		
- PointCloud/Lidar/PolyModels/2DTextures/Meshes/Scene Areas	DRZ	DIE
- 2D Classifier, Metadata Extractor, Wordnet	DNEG	LCM & DIE
WP5		
WP5T1 Asset Transformation Framework		
- Asset representation & transformation options	DNEG	LCM & DIE
- Incorporation of transformation in search results	DNEG	LCM & DIE
WP5T2 Asset editing for re-purposing		
- Crowd trajectory editing	DNEG	DNC
- Animation fusion / splicing	DNEG	DNC
WP5T3 Asset Synthesis		
- Splicing & trajectory tools (Houdini)	DNEG	DNC
WP5T6 Asset Transcode Mechanism (M13-M30)		
- Flix scheduling for conversion, classification etc.	DNEG/FO	DIE
WP6		
WP6T1 Smart Asset use in Virtual Production (M7-24, FA)		
- Character Animation implementation in VP	FA	LCM & LWF
WP6T3 Time, space and world-awareness approach for animation synthesis (M7-24, IK)		
- ML driven based animation synthesis with high-level user commands (world awareness)	UPF (former IK)	LCM & LWF
- Retargeting of animations (to different character)	IK	LCM & DIE
- Animation Stylization	UPF	LCM & DIE
WP6T5 Motion stylization (M1-24)		
- Behaviour tree based autonomous background character animation	UPF	VG
- Emotional stylization of human character animation	UPF	VG
WP6T6 Crowd Scene Synthesis (M1-30)		
- Semantic crowd synthesis/retargeting in semantic environment	TCD	DC
WP7		
WP7T1 Digital Asset Store		
- FLIX Asset Management	FO	DIE
- Integration with Search & Transformation framework	FO/DNEG	DIE

WP7T2 Digital Asset Geolocation		
- FLIX global collaboration capabilities	FO	DIE

4.2 Covid-19 Impact

Due to the Covid-19 pandemic the original experimental production plans reported on in the 'Interim Experimental Production Evaluation Report' (D8.2) had to be modified. Most of the staff has moved to HomeOffice for a majority of 2020, making collaborative productions almost impossible. Studio facilities were closed. Student and commercial productions have been cancelled or delayed. In addition, productions that were still possible in 2020 had to be shrunk down to a minimum amount of people on site and a minimum amount of time spent together with others. This prevented adding SAUCE evaluations to running productions in the scope initially planned.

At Filmakademie, 2 additional productions have been added to cope with the situation. We had the unique opportunity to include experimental productions into an LED Cave shoot in Mannheim (Germany) - LCM including assets form the "Love and 50 Megatons" student project and add another experimental production in the studios of Filmakademie, also including an LED wall and reusing Assets from the "Woody" student projects - LWF. This replaces the evaluation within the student projects at filmakademie as it was not possible to run them due to the above mentioned reasons. By including the new LED aspects, these experimental productions actually provide even more interesting insights than the originally planned ones.

Some experimental productions planned by DNEG and Foundry were difficult to realize due to the various lockdowns, HomeOffice and other Covid-19 impacts. To still be able to evaluate all the aspects these industry partners carried out, they ran additional internal evaluations (DIE) that are reported on in D8.5 and in paragraph 9 of this deliverable.

5 Unfolding

Exploring the possibilities and challenges of lightfields for movie productions was one aspect to be enlightened by the experimental virtual productions within SAUCE. Next to the research aspect, the intention of the"Unfolding" production was to create a trailer for a classic music festival, targeting a young audience by utilizing new technologies during production. The used light field camera has been built by Saarland University and is first tested under production conditions within the "Unfolding" shoot. FA developed the content frame and prepared a complete previsualization for the final trailer. The scientific aspect was carried out to test the light field camera system and recording technology within an environment close to a real VFX driven production. The scenario reflects a typical VFX setup "Person in front of a green screen". Such a setup enables content creators to separate foreground and background elements and to modify and recompose both in a post-production process. The produced footage was used to demonstrate how light field data can enrich a creative process.

5.1 Scenario

In a white room a cellist sits blurred with her instrument. With the first notes she plays, the sharpness slowly expands. During the cello part, the focus adjusts to the scenery through tilt-shift simulations, focus shifts and T-Stop variations. All those effects are created and animated in post-production by exploiting the possibilities the captured lightfield is offering. The production has been supported by the director of photography (DoP) Matthias Bolliger and the cellist Isabel Gehweiler. The support of an experienced DoP for the production was essential because we wanted to be able to define the needs of creatives and identify "real word" problems and possibilities in handling lightfield cameras and data.

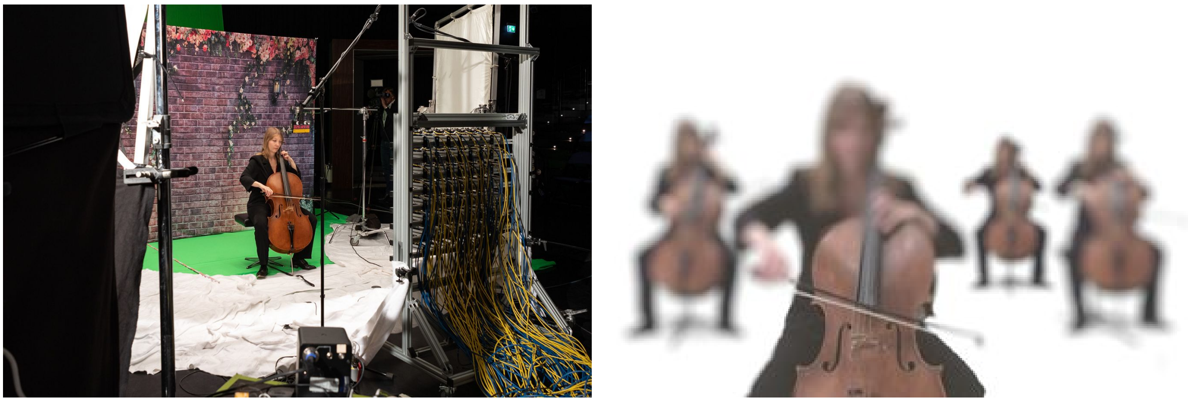
To test the production setup and to provide ground truth data, we designed a 3D virtual version of the scenery. The set has been created in the DCC application Blender, and enabled us to plan the physical setup, stage the camera and define timings. Furthermore we were already able to produce light field data by rendering synthetic images of 64 virtual cameras without the need of a physical setup. This speeded up the whole creative process and gave us the possibility to define the final look. At the same time, it is easier to avoid physical imponderables when handling the complex camera system.

5.2 Recording

The "Unfolding" production was shot in a professional studio at Saarländischer Rundfunk (German Broadcaster). A 5x5 meter greenstage with adequate lighting was realised to create the possibility for color key the footage later in the post production. Furthermore a professional audio capturing and digital recording setup was installed for capturing multi channel audio from the cellists play.

5.3 Post Production

To be able to make use of the captured light field material the RAW data had to be processed. The geometric calibration has been performed with a SLAM++ implementation reported on in D3.3. In this deliverable an in-depth insight into all preprocessing algorithms is available. Two methods for colour calibration, developed by SAUCE partners, have been evaluated as well. For light field camera arrays, a colour equalization step is important to account for differences in colour balance between the different cameras. Two colour equalization schemes are proposed, which use SIFT correspondences between reference and test image pairs to estimate colour correction functions which ensure colour consistency is maintained across all views.



Left: Impression of the 'Unfolding' shoot in the Studios of Saarländischer Rundfunk
Right: Snapshot of the final 'Unfolding 1.0' 2D video

All preprocessing steps (debayering, rectification & colour equalization) were later computed on the cluster being part of the light field camera itself. The 64 computers work jointly to process all frames in a post process after the recording. For the actual post production Filmakademie used a combination of tools developed by SAUCE partners and established production software. A video was generated, showcasing the potential and possibilities of light field video material.
Filmakademie tested whether LF data already offer advantages in existing workflows. In a continuously ongoing comparison the already existing solutions and tools, provided by the project partners were tested against the Nuke based production. Nuke is an industry standard composition software offered by Foundry.

5.4 Recap of Unfolding (1.0)

Goal for the Unfolding production was to utilize the possibilities lightfields offer to enable a VFXartist to separate foreground and background elements without the need of colour keying and to produce realistic, lens simulations of physically possible and impossible lenses.
In order to enable creatives to benefit from LFs possibilities, several challenges have to be overcome. First of all, more size and speed efficient compression methods are required. This would be a starting point to make such data usable in a post-production pipeline. 64 raw video feeds are captured on set by the light field camera array. This enormous amount of data needs to be run through a carefully designed pipeline to be able to get compelling results. As a first step after capturing, slight inaccuracies of the material need to be eliminated. This involves consistent colour equalization across all cameras and a geometric rectification.
Looking at potential benefits light field data can provide, the most useful would be to generate high quality, dynamic depth maps or even geometric reconstructions out of a 5D light field. This would not only help to overcome the need of colour keying but also unlock possibilities that could lead to a workflow to freely edit the captured scenery including its geometric information, lighting and surface characteristics after the actual recording. More defined requirements can be directly derived from the test production. Next to the separation of foreground and background elements, the production would require solutions to change and simulate camera lens effects in Post. Next to this, solutions for moving the camera and changing the camera's perspective after the recording would be very useful tools and could help to overcome the physical limitations of the LF-camera rig. Another aspect is the potential "direct" presentation of the captured footage by a system capable of displaying lightfield data as showcased in the follow-up production Unfolding 2.0.
Further information on this production can be found in D8.2.

5.5 Unfolding 2.0

5.5.1 Concepts

The output of the original "Unfolding" production was a compositing of the lightfield footage in front of a white background, which demonstrated some advantages of lightfields like refocus and tilt-shift effects. For "Unfolding 2.0" we decided to go for more elaborate results like displaying the footage on the "Looking Glass[1]" holographic display and replacing the background with an immersive 3D environment. The goal is to further showcase the versatile nature of lightfield footage.
To achieve this several Nuke scripts need to be developed to help with the automatic processing of the lightfield footage. This includes preprocessing like chroma key, retiming and rectification as well as calculating depth maps for the post production. Additionally we need new or adapted plug-ins for Unity to display our footage on the "Looking Glass".

5.5.2 Realisation

The "Looking Glass" comes with its own format for displaying Lightfield footage. It's called "Quilt" and consists of 45 images arranged in a specific order on a contact sheet. Since we only had the one row consisting of 8 cameras the best way to get to 45 images seemed to be to interpolate some frames in between, we achieved this with the Nuke Kronos plugin.
After we got our first Quilt-footage on the "Looking Glass" we realized that the disparity between the original Lightfield camera array was way to high to be looked at comfortably on the holographic display. It caused some eyestrain and most of the scene seemed to be out of focus. So we decided to go for a realtime 3D scene in Unity instead. This gave us way more flexibility in setting up the camera and world in a way to achieve the best visual experience on the "Looking Glass".

Left: light field frame containing the 64 camera views
Right: snapshot from the 'Unfolding 2.0' holographic output

A quilt is a 4k Image, containing 5x9 smaller sub images. One for each of the required 45 views of the holo display. The quilts are defined as static images, so we had to modify the existing plugin to support video. We also had to add support for transparencies, as this was needed for the integration of the individual cellists into the 3d scenery. To make this possible the HPV [2](High Performance Video) player was used. The HPV Eco-system was originally created within the former Dreamspace[3] Project, funded by the European Union under the FP7 programme. The HPV is a cross-platform C++

[1] https://lookingglassfactory.com
[2] https://github.com/vjacob/ofxHPVPlayer
[3] dreamspaceproject.eu

toolset for playing back high resolution and high frame rate video content and also supports alpha masks.

5.5.3 Results & best-practices

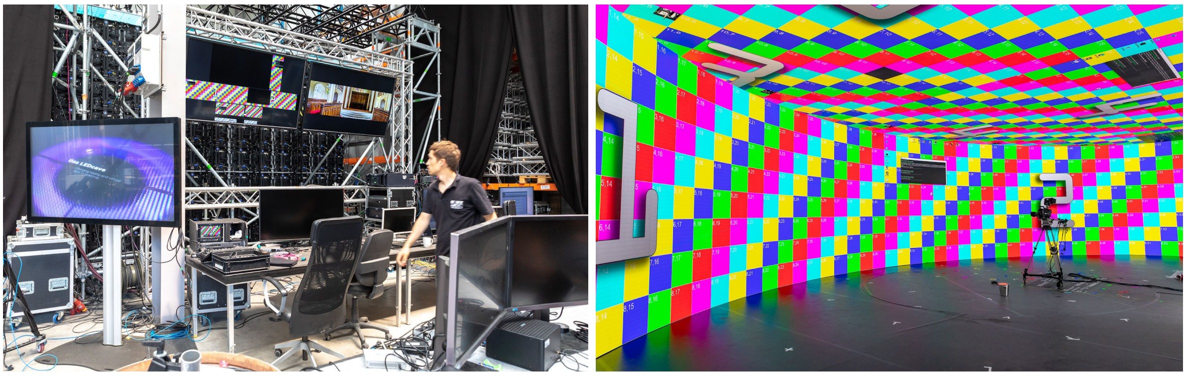
Going for the Unity scene with our protagonist as video texture on cards unfortunately meant that for this current first version we did not yet use the advantages of the lightfield footage. While the scene works very well on the "Looking Glass", our protagonist itself is just flat for now.
For the next iteration we are planning to try two approaches:
1) using the high-fidelity depth of the Quilt-format for our protagonist in combination with the real-time 3D scene in the background
2) using disparity maps created from the lightfield footage to give our protagonist depth with a displacement shader
Material from the initial LF elements shoot has been released publicly and is available as download from our website[4]. It contains raw and cleaned lightfield data captured with the lightfield rig built by University of Saarland. The capturing was organized by Filmakademie and took place in their studio facilities.

[4] https://animationsinstitut.de/en/research/projects/sauce

6 LED Wall Productions

6.1 LED Cave Mannheim

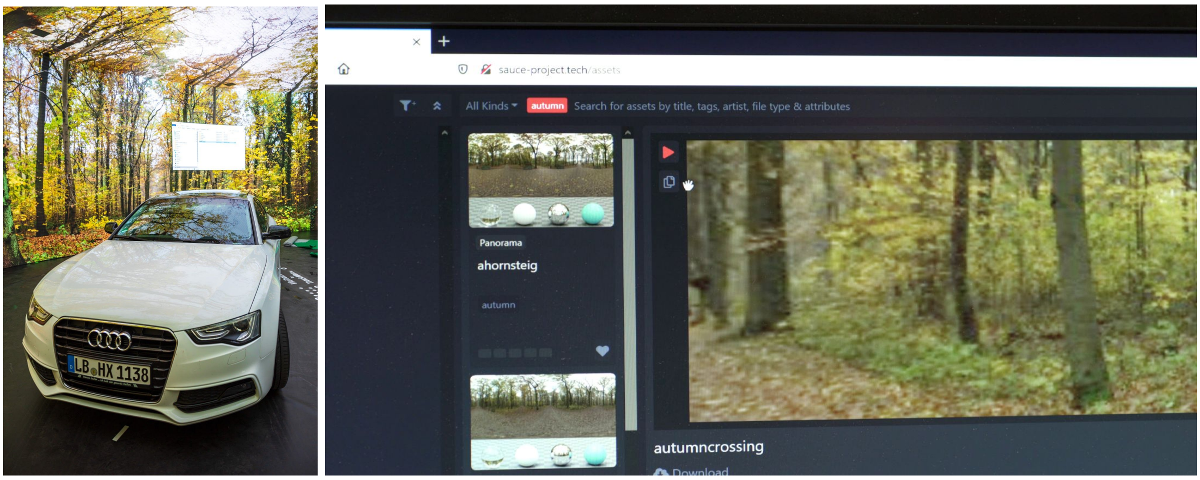
In the LED Cave built by RentEventTec in Mannheim we ran another experimental production.
The cave is a D shaped construction of 3 large screens (front, back, top) with an overall size of 18m by 9.5m and a height of 4m. The curved front screen consists of 464 LED modules with 1.9mm pixel pitch (pp) and has a resolution of 14.848 x 2.048 pixels. The top screen uses 594 Modules also with 1.9mm pp but a reduced resolution of 9.216 x 4.864 pixels. For the back screen 288 brighter, but also less dense modules with 3.9mm pp and an overall resolution of 9.216 x 4.864 pixel were used. 7 synchronized workstations are needed to fill this amount of pixels in a sufficient manner. The render nodes run either Unreal or Unity to generate the displayed sceneries. In order for the recorded image to look correct in the camera, it must be distorted due to the camera position and rotation. For this purpose, it is necessary to be able to determine this data from the camere at any time. A 12 Camera Optitrack Prime41 Motion Capturing System with passive marker was installed in the cave.
In order to preserve the light characteristics created by the cave and at the same time reduce image artefacts like moire, a professional cinema camera and fast lenses (Sony Venice, Zeiss Ultra Prime Lenses) were used during the shoot.



Impressions from the LED Cave in Mannheim: control center (left), led stage (right)

We were able to test the pipeline of quickly exchanging background and lighting in a car shoot scenario. The assets were searched and selected within the search and transformation framework developed by DNEG and then displayed in the LED cave. Additionally we ran first tests with a machine learning driven character animation provided by UPF that was directed again in real-time on our VPET tablet clients.

6.1.1 Search and Transformation Framework in Carshoot Scenario



Right: The search and transformation framework in use at the LED Cave
Left: The selected HDR 360° image on the LED stage

One big advantage of using an LED Wall or Cave in a movie production is, that background but especially also lighting is provided by it. Additionally the imagery displayed on the LED screen can be exchanged quickly. This quick exchange also requires tools to easily search and find appropriate content.

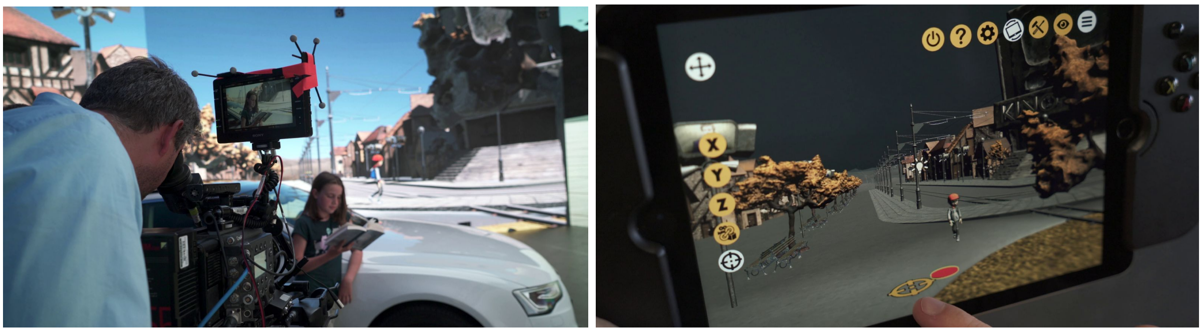
Imagine for example a director of photography, set dresser, creative director etc. requesting a change of the background and lighting. While technically quickly possible the content also needs to be generated or reused from a database. The search and transformation framework developed by DNEG within SAUCE can provide this. A database of assets, for the LED Cave experimental production 360 HDR images, can be searched for while the assets are stored along with searchable descriptions and tags.

Within moments it was possible to search e.g. for an autumn forest in the search frontend, extract the image from it and display it on the LED cave, changing the entire scenery and lighting. The Foundry's FLIX backend provided the networking architecture for this. Assets were searchable in a web browser at the venue of the LED wall in Mannheim (Germany), while the actual asset could be stored anywhere, e.g. in the DNEG headquarter in London.

The Search and Transformation framework also provides the option to add transformation plugins which makes it even more valuable when reusing assets from previous productions using other data formats, resolutions or panoramic shapes.

The search framework has been developed within WP4 (WP4T1, WP4T2, WP4T3), the transformation plugin architecture has been added in WP5T1. The flix backend and integration with the S&T framework was developed in WP7T1.

6.1.2 Character Animation in VPET



Character Animation in use: Professional DoP Matthias Bolliger at work on our set (left), VPET tablet client for real-time character control (right)

Virtual productions get increasingly common in modern movie productions. The possibilities to visualize, edit and explore virtual 3D content directly on a movie set make it invaluable for VFX rich productions. Many of the virtual production scenarios also involve animated characters and motion capturing. But the complexity of animation systems prohibits it's usage on a film set. While most of the virtual production tools and frameworks are not publicly available or open source, none of them has the possibility to interactively and intuitively animate characters on set.

Besides other topics, approaches are developed to make character animation and the involved characters themselves 'smarter'. This involves procedural character animation and machine learning being used to provide high level control over a character.The R&D team of Filmakademie Baden-Württemberg developed the open source 'Virtual Production Editing Tools' (VPET) over the last years. With as little hardware overhead as possible, VPET offers the possibility to stream an arbitrary 3D scene to tablet clients. On the tablets, the 3D scene can be aligned with the real world in augmented reality (AR) making the tablet a window to the provided set extension. Users can explore and edit 3D elements, lighting as well as rigid body animations. All clients and the scene host communicate changes among themselves through a synchronisation server, keeping the scene consistent.

To approach character animation in the virtual production toolset FA developed an open character streaming protocol for VPET. The entire character (including weights, skeleton etc.) can be transferred to the tablets at run time. The newly developed API then allows arbitrary external animation solving engines to animate the character through streamed bone animations. These animations are represented as a root bone translation and a new rotation for each bone. The updated pose is automatically synchronized and held consistent between participating tablet clients.

It does not make sense to author a complete animation from scratch with an on-set virtual production tool set, as it requires to much time and expertise to get a convincing result. Nevertheless providing the possibility to easily direct virtual characters on a film set is often desired. Tablets offer an intuitive way of interacting with elements during a virtual production e.g. in augmented reality (AR) being useful for directing characters. Flexibility and ease of use are the main targets for our work.

Towards this goal, FA and UPF have been working on a joint effort within the SAUCE project. On set, only high level commands can be used to drive a character. Commands like 'Go there', 'Run' etc. should be used on e.g. the VPET tablet tools. This requires that procedural animations are generated and that the character is animated in a scene aware manner. Obstacles should be avoided, uneven grounds need to be compensated etc. This technology is also known from game creation, but only slowly they are introduced into the film industry. Such complex, highly adaptable animation solving, required for virtual productions, cannot be executed on a tablet conveniently and with the required customizability.

For the LEDCave experimental production a machine learning implementation developed by UPF and formerly IKinema has been used as an animation engine plugged into VPET. Through the core technology, the learned animation can be applied to an arbitrary human character. This approach is thought for virtual characters which receive the orders directly from the artist. The user interface for character control in VPET and the usability and usefulness in a virtual production environment have been successfully tested in this experimental production.
UPF also worked on another approach, focused on automatically controlling the background characters, reducing the time spent if this task was done manually. UPF has developed a web tool in order to determine the characters' identity, including the behaviour and the style. This background character will be controlled by a graph programmed Hybrid Behaviour Tree.
Through the introduced VPET animation streaming API, both approaches are easily integrated into the tablet tools and can even coexist within the same scene. FA provided the training data for the ML algorithm as high quality optical motion capturings. This database is also released publicly.
All in all the combination of a new character animation streaming API and two animation solving engines for different purposes form a good foundation for adding intuitive, high-level character animation to an on-set virtual production.

6.1.3 Evaluation and best-practise

To be able to make the best use of an LED wall setup, the demand is high to have fast and intuitive tools available on set to find and add assets as well as direct and edit them.
VPET is developed to edit and explore 3D sceneries on set and therefore already proved to be an essential tool is such a LED wall based virtual production scenario. With the added character animation interface, VPET becomes even more valuable and makes it possible to also direct virtual characters intuitively and on a high-level in real-time on set. Especially for the previz production phase but also in the final production, this can become very handy.
The developed S&T framework and the FLIX backend provide an essential architecture to stay agile and fast in a demanding virtual production process. Reusable assets can be found and added to the scenery quickly.

6.2 LED Wall at Filmakademie (Snowflower production)

For evaluating the contributions of all project partners in a final state and testing if and how well the technologies can be used in combination, Animationsinstitut of Filmakademie ventured a final showcase. To avoid a mere handling of test sequences the team came up with a small but complete production scenario, including story development, asset preparation and a two-day studio shoot. Under conditions close to reality this experiment would allow for robust findings while also producing interesting footage for further research and documentation.

6.2.1 Concept

The main focus was on character animation driven by machine learning and its control with editing tools directly on set. The real-time nature of these techniques demanded a virtual production environment, featuring real-time camera tracking, rendering and displaying, in this case on a large-scale LED wall. Related to that it would also be interesting to examine if screened characters can be used in combination with live action and to what extent they can even interact with the foreground.

6.2.2 Preparation

6.2.2.1 Story Constraints

The usage of machine learning for character animation limits the range of motion to the provided training material, while the tracking volume of an optical motion capture system as well as the dimensions of the room and the specifications of the LED display crop the range of motion of the camera. Hence, there were some constraints to the possible movements and thus to the story.

The characters were capable of walking and running in all directions, could step backwards and sidewards and knew how to pass over obstacles. The camera was allowed to move in a space of approximately 8 by 14 meters, not getting too close to the LED wall as camera tracking there suffered from occlusion and also not moving too far away to ensure that the display was frame filling. Additionally the LED wall, i.e. the CG background, should not be in focus to avoid moiré-artefacts.

6.2.2.2 Story

Considering the described constraints and the fact that the purpose of the production was to allow for an in-depth evaluation of techniques, the team kept things simple when working on the story.

The first shot shows a sparsely furnished room which is lit by a flickering fireplace and some faint lampions. Through the opening of a broken door a winter landscape is visible. A vitreous vase with a single flower stands on a small table in the middle of the room. Suddenly a humanoid ape-like creature appears in the distance, approaching slowly. It climbs through the broken door and enters the room, cowardly looking around. The camera jumps to the vase and captures the refraction of the ape getting closer. The creature grabs the flower, twists around and rushes back into the woods.

The short film was designed to demonstrate how the animation system is capable of reacting to obstacles (wall with door) and illustrated the influence of different areas (snowy outside world, sinister room) on the body language of the creature. Also the need for controllability of such synthesized animations is apparent. The provided editing tools would enable the team on set to manipulate not only the path and timing of the movement but of all kinds of virtual set objects as well.

6.2.2.3 Asset Preparation

The room itself as well as all interior objects were built in Cinema 4D, further prepared in Maya and then textured in Substance Painter. Unity would be the backbone of the production, since the machine learning framework had been set up here. Also Unity would handle the communication with the tablet clients to allow for animation control and set editing. Consequently shading, lighting and further scene preparation were done in the game engine. The winter landscape was bought in the Unity asset store. For every object in the scene, a duplicate had to be created with lower resolution IOS-ready textures to be streamed to the tablet clients.

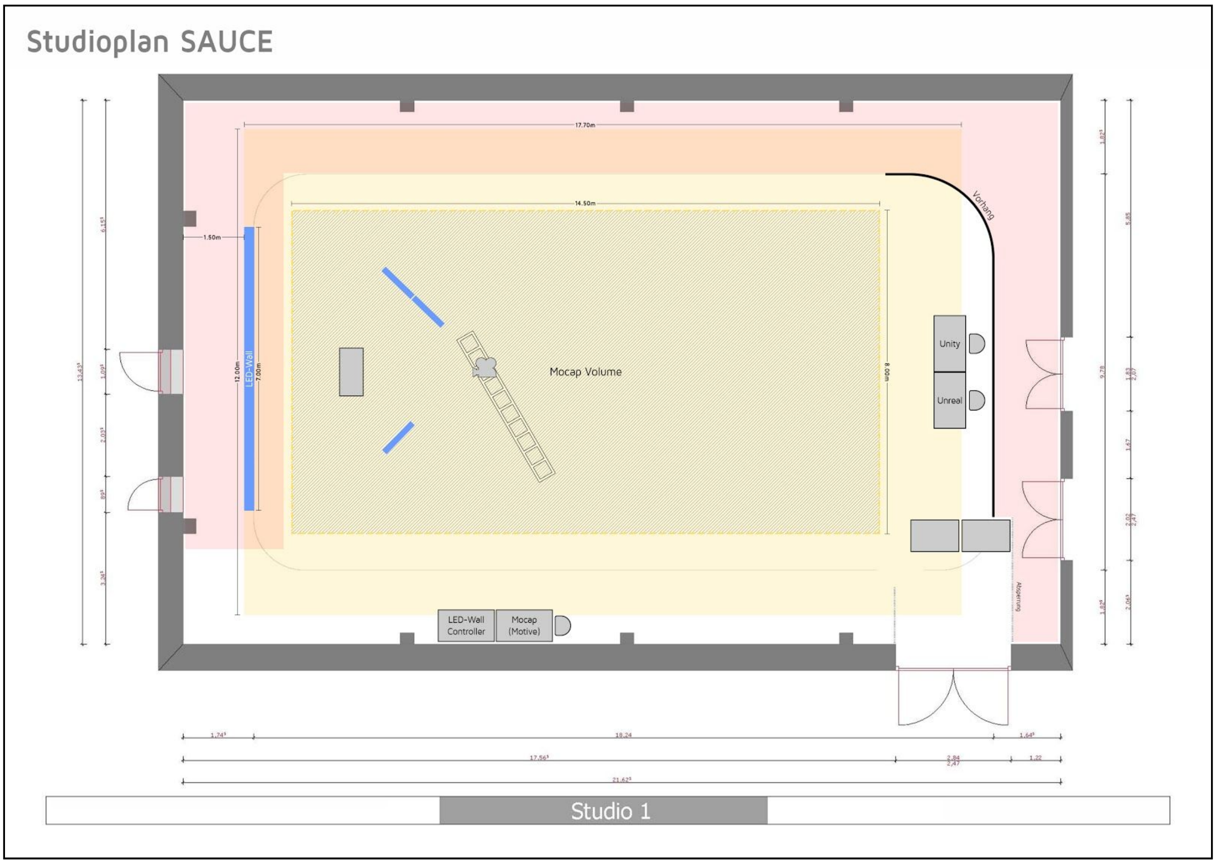
The ape creature had originally been built for a student production at Animationsinstitut using Houdini and heavy simulation. In order to be able to use the character in the context of a game engine it had to be simplified, re-textured and re-rigged. As a backup the character was also exported with baked mocap animation out of Motionbuilder.

Rendering of the virtual interior scene used in the Snowflower production

6.2.3 Technical Framework

The studio setup encompassed a flat LED wall with a width of 7 meters and a height of 4 meters, consisting of 14 by 8 individual panels with a pixel pitch of 1.9 mm, resulting in a total resolution of 3584 by 2048 pixels. As maximal brightness 1200 nit are stated[5]. Three additional panels with a size of 0.8 by 1.8 meters, a pixel pitch of 8.33 mm and a maximal brightness of 7000 nit[6] got mounted on wheels and were mainly used as movable light sources or for casting reflections. The entire LED setup was rented from the company Rent Event Tec GmbH for an academic studio workshop and not provided for SAUCE exclusively. The camera and the three light panels were tracked live in the studio. For this purpose an Optitrack motion capturing system with passive markers was used. The markers were placed on top of the camera and on the three moveable panels. In order to cover the entire area, 24 motion capturing cameras were installed on all four walls of the studio.This allowed us to determine the exact position and orientation of the camera and the light panels at all times.
The acquired data was transmitted via network to a workstation running Unity and the VPET server and infrastructure.

[5] https://www.leditgo.de/files/pdf/LEDitgo_rXone_Datenblatt.pdf
[6] https://www.leditgo.de/files/pdf/LEDitgo_sB8_Datenblatt.pdf



Studio plan of the used setup (LED walls in blue colour)

To avoid artefacts in the animations and the captured image, the rendering, animation, camera and LED wall had to be synchronised. The systems used offered such a possibility. They were connected to an external genlock, which provides a world clock of 50 frames per second.
Despite the comparatively bright LED technology, it is still an advantage to use particularly light sensitive cameras and bright lenses. This helps to make creative use of the lighting provided by LEDs. Due to the low depth of field such a setup also helps to maintain a cinematographic look ands prevents image artefacts such as moire. Therefore, a Sony Venice[7] Camera in combination with Leica Summichon C[8] lenses was chosen as the main camera. To underline the soft and warm scenery and to further reduce resolution related artefacts an additional softening filter was installed in front of the camera. For the interaction with the scenery Apple iPads with VPET clients were used, which were connected to the VPET server workstation via a Wifi network.

[7] https://pro.sony/de_DE/products/digital-cinema-cameras/venice
[8] https://www.leitz-cine.com/summicron-c-premium-lenses-for-larger-sensors-film-and-television-productions

Studio production photos

6.2.4 Shoot

For the shoot the professional DoP Julia Schlingmann supported camera setup and cinematography, also giving essential feedback on the virtual production environment in general as well as on the quality of the content displayed on the LED wall in detail. Like this the research team could concentrate on preparing and using the character animation framework and the related control techniques.

The first shot established the vase on the table with a slow forward track, gradually revealing the creature approaching in the distance. Since the first shot would serve as reference for all three shots in the short film, it was thoroughly arranged, playing with different lenses and LED levels and paying attention to smallest reflections. In multiple run-throughs the ape got animated using both pre-recorded motion capturing and machine learning, the latter requiring the input of the tablet-based editing tools developed by Animationsinstitut.

The second shot showed the ape only as reflection on the water surface, while the camera moved gently to induce some parallax. The grabbing of the flower was held out of frame, the audience would only see the stem being pulled out of the vase. Here a precise positioning of both the ape and the environment was needed to create the desired look.

The third shot picked up the framing of the first one and, by slowly tracking back, followed the creature as it strode away into the forest.

For all shots a video of a flickering fireplace was screened on two of the three movable LED panels and provided natural refractions and reflections, especially on the vase, which was furthermore held in focus to avoid moiré. Apart from one extra back light, the LED wall and the movable panels were the only light source and sufficient for illuminating the foreground. A black pro-mist filter was used to soften lights and contrasts. This halation helped to further harmonize foreground and virtual background. By adding some depth blur to the images rendered by the game engine the overall depth of field looked slightly more convincing, although simply adding virtual defocus and real bokehs usually does not end up in realistic results.

In general the idea of screening characters on a LED wall and combining them with live action foregrounds proved to be feasible with huge limitations. As soon as the character gets too prominent in the image, one would expect it to be in focus, which is not possible since the LED panels should stay de-focussed to avoid moiré. Implicitly a real interaction between virtual characters and live action foreground is not possible. Also the fact that the LED wall is in a fixed distance to the camera and thus all screened content holds a constant level of bokeh further detracts from the credibility of the

captured images. Our use case was constructed around these limitations and therefore produced acceptable results. For real productions such screened virtual characters will only be suitable sporadically, for example for background crowds. Future VFX productions could make use of characters via LED screens to engage all actors in the scenery with virtual protagonists. Not being able to focus on the screen and character could be solved by an additional compositing step.
As virtual production processes can be applied not just to VFX productions but also to full CG productions the proposed workflow could lead to completely new production processes. Sceneries could be prepared accordingly and animations be blocked instantly according to the directors expectations. Animators could refine this data in a post process. This proposed workflow would combine Motion Capture, Machine Learning powered tools and hand animation.

6.2.5 Evaluated Aspects

6.2.5.1 Character Animation VPET

Character animation on the VPET tablet clients on the LED wall set of the Snowflower experimental production

One aspect the FA R&D team has been developing within SAUCE is the on-set character control with our virtual production editing tools. And we are testing this now in our experimental LED wall production. A user can use our virtual production editing tools called VPET not only to intuitively direct light and objects, but also characters. This can be done by selecting it in the scene on the tablet client and using the joysticks to control the movement of the character directly on set without operating it from a desktop computer. The animation is solved through machine learning. The

neuronal network generating the animation is running on the computer rendering the scene. The resulting animation is displayed on the LED wall and streamed back to the VPET clients. The character is then also able to interact and react to the virtual scene surrounding it.
The machine learning animation learns the walking of for example a human character from motion capture data. We captured a database of 1 to 2 hours of different movements that are fed into the training network where it takes 16 hours to build a model. Afterwards it can generate the animations based on the user input on the VPET tablets in real-time.
The produced motion capturing database is released publicly on our website[9].

6.2.5.2 Mesh reduction transformation

The wooden ape character called Woody has been produced by a student project here at FA. We reused this asset in our LED wall production. To be feasible for the used game engine, mesh, textures and rig had to be reduced in complexity. Actually all assets one can see on the LED wall had to be prepared for this realtime rendering application.
Additionally a reduced version needs to be provided to the VPET tablet clients. The Search and transformation framework developed by DNEG can be used to find reusable assets from past productions. Additionally it provides an API to automatise transformations like the mentioned mesh reductions.

6.2.5.3 Mood region Classification

Disney Research Studios Zurich developed machine learning tools to classify entire virtual sceneries. We called them mood regions. The implementation is able to identify which mood a specific region is likely to cause.
In our LED wall scene, the snowy outside region and the indoor region will cause different emotions or moods of the virtual ape character. By automatically labeling these regions, an animation engine could adjust the animation according to the scenery surrounding the animated character, fully automatic. This reduces the complexity of the on set character editing.
In addition DRS developed deep learning networks to identify all kinds of 3D assets. The thereby generated labels can be consumed by for example the search and transformation framework developed by DNEG. Assets are thereby made smart and reusable for future productions.

6.2.5.4 Crowds

Not only a single character can be animated on such an LED wall. The Trinity College Dublin also developed a crowd tool for the Unity game engine. It is also based on machine learning and allows crowds to react on the surrounding character and scene objects, while also providing the possibility to be directed by an artist.

6.2.6 Results & best-practices

The production showed machine learning, smart assets and interactive interfaces are key to be able to run virtual productions especially when the goal is to capture 'everything in camera' by e.g. using an LED wall. This reduces or eliminates the needs for an extensive postproduction but adds high demands on pre-production.
All virtual assets need to be available in final quality before the shoot. Any adjustments to the scene need to be done on set in parallel to the actual production. The search and transformation framework including it's transformation capabilities can become a very helpful tool in this process as it allows artists to quickly add new, reusable assets matching the requirements of the production. The FLIX asset reuse backend makes it possible to collaboratively source these assets from storage facilities all over the world.

[9] https://animationsinstitut.de/en/research/projects/sauce

More and more virtual characters are used in modern productions. In a virtual production scenario they need to be directable on set just like an actor. This can only be made possible by providing tools to direct the virtual characters with high-level commands to the persons incharge on set. The character animation API added to VPET was made exactly for this purpose. The flexibility of choosing an appropriate animation engine (e.g. a machine learning approach incl. Mood regions, crowd tools) provides the ability to adjust to a variety of production scenarios which becomes very valuable since VPET is tailored more and more towards a general purpose real-time editing tool for 3D content and separates it from other available tools being reported on in D6.2.

7 UPF Video Game Production

The experimental production carried out by UPF-GTI has the purpose of validating the abstraction and the integrability of the libraries developed for the WP6T5, as well as to prove the usability of Medusa's output outside the standalone tool. To do so, we implemented a proof of concept of a video game on the web which uses all the features developed inside Medusa. Furthermore, we carried out a user evaluation experiment.

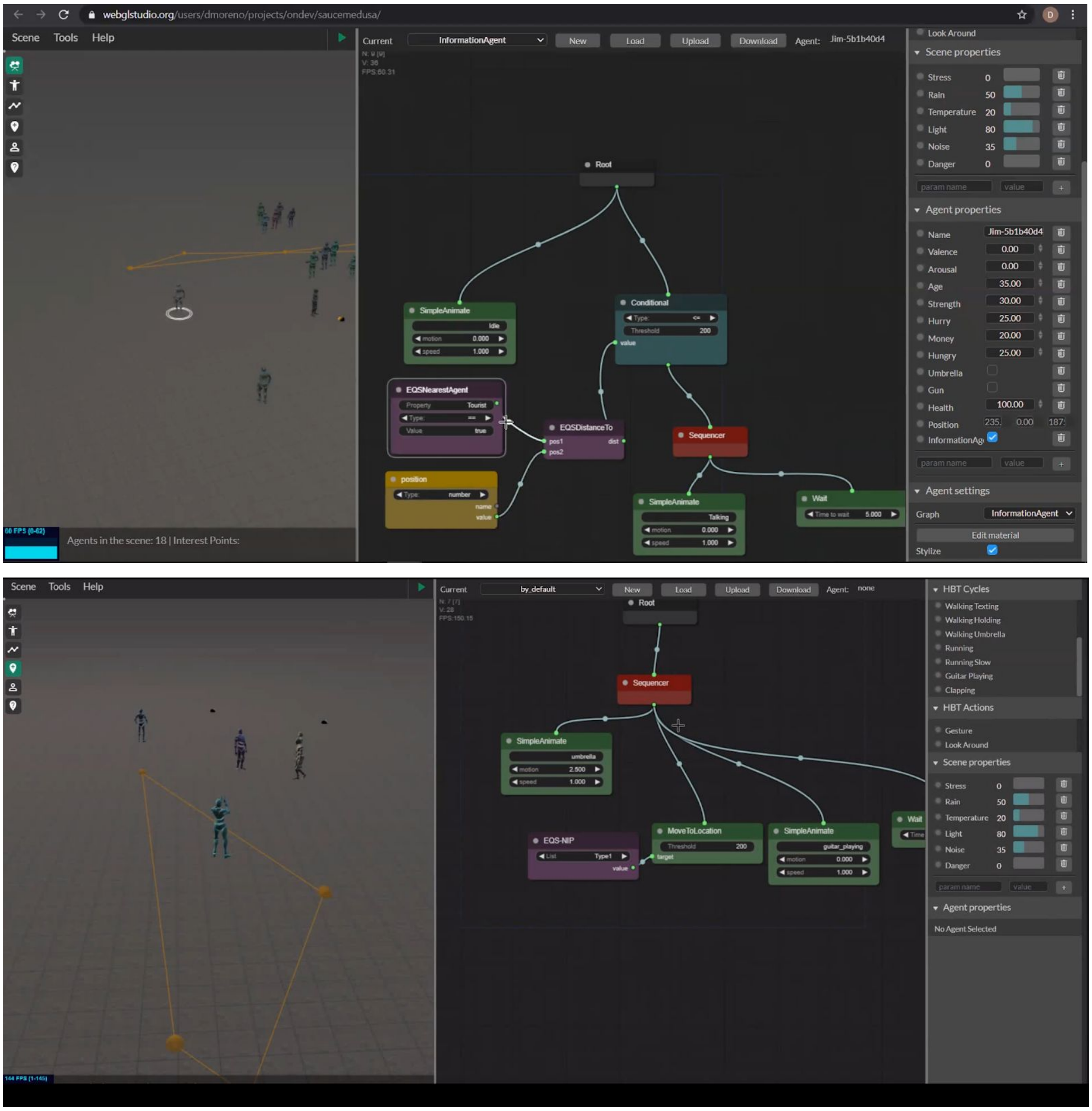
7.1 User evaluation experiment

In order to evaluate the standalone development, we prepared an experiment that was performed by different person profiles, including students, freelancers working on Unity with some plugins published, Ubisoft game programmers as well as researchers on the field. We really appreciated their collaboration in this user evaluation experiment. All of them had to perform the same tasks, which included:

- Create basic behaviours for virtual characters without having into account their environment, such as behaviour conditioned only on their inner properties.
- Create a little more complex behaviours where points of interest and another virtual character were taken into account to perform the final goal of the task.
- All of the previous tasks included achieving different animation styles through the emotional stylization of the virtual characters

All of the evaluations were completed successfully, as we consider that completing all of them was a successful result of the experiment. Despite that, some help was required during the development of the tasks. Most of the doubts appeared due to some interactive confusions with the tool, as well as some misunderstanding of the performance of the most complex behaviour tree nodes, such as linking several *Sequencer Nodes,* which leads to slightly more complicated comprehension of the tree execution. It is remarkable to mention that the guide and videos provided prior to the experiment resulted quite useful for the users, as they personally claimed, but it has been improving continuously taking into account their concerns and doubts during the experiment. Also it is important to mention that the expected duration of the experiment was between 35-45 minutes, and the average of the experiment's durations is 46 minutes, not so far from our initial expectations.

The overall conclusions of the user evaluation experiment, taking into account the performance of users with no experience in Medusa (neither graph programming nor behaviour trees in some cases) and their final comments and suggestions, were positive. They found the tool and the stylization system highly intuitive and, despite some minor doubts, they found it useful to have such an easy to use system for fast prototyping of background environments with animated semi-autonomous virtual characters. Also, almost all of the users claimed that, improving the tool interactivity (taking into account their own feelings and their own suggestions) would make it even easier and more usable for a wider audience.



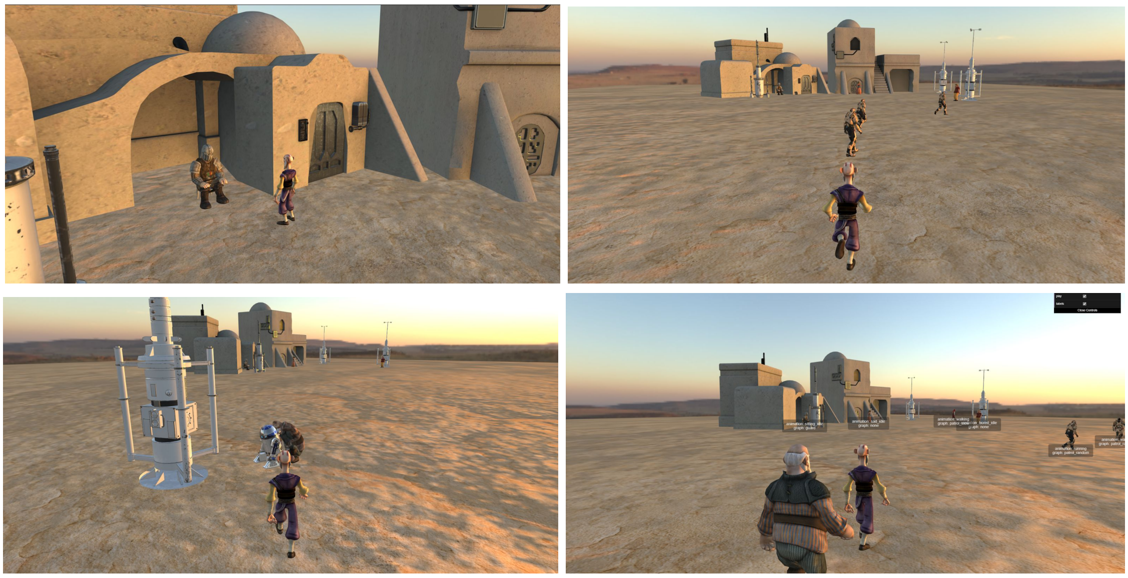
Some screenshots of different user evaluation experiments using Medusa

These user evaluations helped us to improve and validate the implementations carried out in Medusa, allowing us to upgrade the main strengths we aimed to achieve with the tool, such as integrability, user interaction and understandability. Moreover, it was beneficial to have a wider perspective on what could be done using Medusa and also considering new features or applications.

7.2 Proof of concept of video game

Regarding the proof of concept of video game development, as mentioned and planned in the D8.2, we aimed to integrate the full behaviour control system outside Medusa, and evaluating from basic to complex NPCs behaviours, covering a wide range of different things take in into account when computing the behaviour, such as environment, other NPCs, or even the main player. This videogame proof of concept is not conceived as a production videogame, but it aims to showcase the basis of what could be done using our developments.

For example, we have designed a use case where the main player has the mission to guide a space mechanic to find a right place, but in case the mechanic loses contact with the main player when moving to its desired location, gets confused and lost. Within this simple example, Medusa core is taking into account the environment, the main player dynamic behaviour (controlled by the user), as well as inner properties.



Screenshots of the proof of concept of video game

8 TCD DublinCrowds

The purpose of this experimental production was to evaluate the semantic crowd synthesis and retargeting tools developed in task WP6T6, which are described in detail in D6.8. Due to the pandemic, the experiment was performed by freelance artists using a newly developed system, which integrated the novel crowd tools described in D6.8 in Unity.

The task was to populate given semantic environments by crowds with believable behaviour. Performance measures were automatically recorded, and dedicated questionnaires had to be filled by participants. Participants were able to successfully create a crowd scene with little knowledge of crowd simulation and we anticipate that our tool can enable this as a creative possibility for small studios. Based on these data, a formal HCI study was conducted to evaluate performance in detail. The results of the study as well as design, implementation, execution, etc. are reported in detail in D8.5. We also intend to publish the study in a scientific paper. The following images show screenshots of example crowd scenes created using the crowd simulation software.

Example crowd scenes created using the TCD crowd simulation software

9 DNEG Crowds

To evaluate the work done for deliverables D5.4 and D5.5, DNEG employed the time of a senior crowd artist to devise a series of tests that represent typical tasks for crowd shots. The test was comprised of three scenarios with three artists, which was able to give us a reasonable snapshot of how useful the work has been.
The tests covered a range of use cases, testing the following scenarios.

9.1 Collision Avoidance

Whilst collision avoidance is common, mostly it relies on a simulation based approach. The new tooling allows for path deformation to be applied after the crowd has been created. The tests involved cases where the agents must avoid each other, and also static objects.
Testing found that the biggest advantage of this approach is the fact that the locomotion resolution is independent from the partial blending. As a result, the work in an artist’s scene is effectively split into two stages, and once happy with the overall crowd simulation, the artist can do passes on the additive blending without having to go back to simulating the locomotion as they would usually do without the SAUCE tool.

9.2 Sharp Turns

The sharper the turn a character has to make, the more necessary it is for appropriate animation to be transitioned into rather than simply deforming the character's path. The tooling is designed to attempt to automatically transition into appropriate animation, so should speed up the turnaround of shots for artists.
Testing found that the biggest advantage of the SAUCE tooling is that once motion is set up by the artist, populating the scene only requires drawing the path where agents are necessary. In previous approaches without the use of the SAUCE tool, artists would need to manually pick clips and offsets individually for every placed source point; the SAUCE tool shields the user from the necessity of doing this.

9.3 Stacked Animation

Sometimes, an artist wants full control over a sequence of animation that needs to be applied to an agent in a specific order. This is doable using common tooling, however, unless the animation is specifically designed to play one after another, artists will experience 'popping' of animation. The new tooling allows artists to stack these animations with automatically generated transitions to ameliorate these issues.
Testing found that this solution would likely scale much better when faced with more than two states, as was tested, and for locomotive states transitioning to other locomotive states with different speeds.

sauce

10 Video documentation

This chapter collects all the documenting videos for the experimental productions as well as recordings of presentations involving them:

Video	URL
Lightfield Elements Shoot documentation	https://youtu.be/aBCKIC_Mt6A
Lightfield 'Unfolding' Shoot documentation	https://youtu.be/5eGWY5RS530
Unfolding 1.0 clip	https://youtu.be/UnsmKQjO4ro
SAUCE lightfield Talk at FMX 2019	https://youtu.be/SaYwLTyvahg
SAUCE TechTalk Stuttgart	https://youtu.be/2z1zJBbiM7g
LEDCave Mannheim Teaser	https://youtu.be/fOEQ26VLGP0
LEDCave Mannheim full coverage	https://youtu.be/kT70eAbA_QY
SAUCE Asset Libraries Talk at DigiPro 2020	https://youtu.be/YeND9dWc2jw
VPET Unreal Engine integration documentation	https://youtu.be/ROlDiYnoEPc
SAUCE Final Demonstrator at FMX Spotlights 2020	https://youtu.be/MVYAHetjUoY
SAUCE Trailer	to be released soon
Snowflower clip	to be released soon
LED Wall experimental production at FA	to be released soon
Further videos will be released on the FA's R&D youtube channel: https://www.youtube.com/RDFilmakademie	

11 Conclusion

The experimental productions that have been carried out showed, that there is high demand for technology and solutions that have been developed within SAUCE. Especially when it comes to movie productions utilizing a variety of computer aided technologies like real-time rendering, motion capturing, camera tracking etc. tools are required to direct, adjust and interact with the virtual content fast, intuitively and reliably.
The trend of moving away from classical productions using green screens to the usage of LED walls and real-time rendering opens new worlds for the production but also puts a lot of expectations on the technologies being used. The character and crowd animation tools, the search and transformation framework and the machine learning based asset classification and the virtual production editing tools (VPET) are perfect examples on how to address these new requirements.
Lightfields adds a significant amount of work during capturing, but the result is a reusable smart asset that can be used and reused in various ways in post-productions of existing and future productions as our Unfolding productions showed. Lightfields capture significantly more of the recorded performance and provide the foundation also for algorithms and software yet to be developed. The process of exploiting these benefits in production pipelines only just started.
Character animation and Crowds tools like the ones being developed within SAUCE can not only be used in movie productions. Also games can greatly benefit from these flexible, user directable animation tools as UPF's video game experimental production showed.
A variety of videos of videos has been produced by FA to document the different experimental productions.