Abstract: This document presents a novel representation for digital 3D scene content. The representation is motivated by advances in realistic scene content renderings, which allow moving content distorting effects from acquisition time to the post-processing stage. At the same time acquisition decisions can be modified in later processing stages. This features are enabled by a representation which stores not only visual content of a scene, but a multitude of additional information. As important as the representation is the presentation of the scene content. This is done by a renderer designed to fulfill the requirements this novel way of representation poses. Both, representation and renderer, present the core functionality enabling a novel and enhanced multimedia experience.

Keywords: Computational Videography, Realistic Scene Reproduction, Multidimensional Scene Representation

1 INTRODUCTION

SCENE is an on-going research project dedicated to create and deliver richer media experiences [1]. A consortium of international research and industry partners aim to enhance the whole chain of multidimensional media production. These enhancements include new capturing devices, scene content processing tools, renderers dedicated to render SCENE data. At the core of this project is a novel representation architecture. This novel architecture aims to provide as much information for postproduction as possible. With the final goal of synthetically recreating camera effects and using higher quality data for intermediate processing the proposed architecture changes the way image and video data is produced. Therefore SCENE brings a paradigm change in the way of visual media production.

This paper is structured as follows: In the next section the change of paradigm introduced by the SCENE project and the historical motivation for this change are explained. Section 3 explains the fundamental idea of the SCENE concept, which is providing additional and better data for further processing steps. This data needs to be acquired with novel capturing devices and techniques. Section 4 contains the conceptual description of the scene representation architecture, highlighting the features and advantages of such a layout. The paper continues with the actual implementation of the envisioned scene representation. The final section concludes the paper and points to future work.

2 SCENE – A PARADigm CHANGE

Throughout history the bottleneck of image or movie capturing devices has been the film; in recent times the image sensor. As the sensitivity of the film or image sensor was comparably low, this bottleneck enforced constraints on the optical system and the capturing process. For low light conditions long exposure times or large lenses had to be chosen; the first resulting in motion blur of moving objects and the second limiting the depth of field. These artefacts have coined movie productions throughout the last century; they even became desired artistic elements and stylistic devices in movie productions.

During the last years new chip technologies have enhanced available image sensors to a level where this physical bottleneck is largely removed [2]. The amount of light necessary to create an image does usually not dictate camera parameters any more. High-speed cameras can capture several thousand frames per second almost fully avoiding noticeable motion blur, and cameras with pinhole lenses allow almost infinite depth of field. Nevertheless, motion blurs and limited depth of field are still applied for artistic means (see Figure 1).

Figure 1: Limited depth of field and motion blur in the movie “Fast and Furious”

Computational Photography alters image content by computational means to create visually appealing and artistically interesting results [3]. Successful implementation of ideas from computational photography requires high quality data and information on the scene content. The same holds for computational videography, which transfers the ideas of computational photography to motion pictures.

Data distortions introduced for artistic means as described above limit the application of computational videography and therefore limit the artistic freedom in post processing steps. SCENE changes the way data is acquired by striving to capture as much undisturbed information as possible by maintaining artistic freedom and directors decisions. Thus SCENE enables the full spectrum of computational videography, limiting neither director nor camera man in his creative freedom.
3 CAPTURING SCENE CONTENT

Data acquisition in the SCENE context can be arbitrarily complex. Of course video data can be captured by traditional means, such that a scene only contains a simple video. More advanced approaches allow capture of depth or environment information.

As part of the SCENE project a motion scene camera is developed. This camera can capture depth information in addition to colour information, by extending the image sensor by a time-of-flight sensor [4]. Depth information is fundamental for scene reconstruction, to on the one hand correctly insert further objects in a scene, and on the other hand to correctly render scenes with synthetic camera settings. Triangulation is a second approach followed in the SCENE project to obtain depth information. Light field cameras as an additional source of information can add further valuable information which can be used for photorealistic synthetic camera effects. Figure 2 shows these three camera types.

Environment maps can be acquired by 360° spherical imaging systems, like the Ladybug produced by Point Grey. Environment maps provide essential information about light sources which are required to relighten objects correctly and adjust lighting conditions synthetically. [5, 6] Valuable additional information can be contributed by any other sensor equipment. Camera calibration data facilitates scene composition, segmentation maps and coherence information enable object tracking and automatic modification.

In addition to scene content captured in the real world, computer generated objects can be inserted and composited with real scene content. From the virtual world there is no limit to the data that can be used in SCENE.

4 THE SCENE REPRESENTATION

The SRA is a key innovation to enable the paradigm change described in Section 2, and to merge and manage the multitude of data coming from separate sources as described in Section 3. Major achievements are

- Single Format: When processing multidimensional video data on a computer a multitude of information sources are required: Video from several sources, camera calibration data, lighting information and spatial knowledge are just naming a few. Our proposed architecture unites all this information necessary for movie production in a single format.

- Undistorted data: When introducing artistic elements like motion blurs, depth of field or colour offsets these effects traditionally modify the captured data. Post-processing such data is time consuming and difficult. The scene representation stores all data in the best available quality and introduces altering effects in a higher layer, thus preserving all available data for facilitated image and video processing steps.

- Content Interaction: Image or video content is usually frame based. The scene representation is object based and therefore allows segmented content. Knowledge about objects in a scene allows interaction such as updated product placement, object modification or camera interaction.

- Unified Representation: Computer Generated (CG) content and Captured Video (CV) stem from two very different worlds and are processed largely independent in movie productions. The scene representation allows a unified representation of both, CG and CV data as well as any intermediate processing steps, thus merging both worlds in an early stage and facilitating post production.

These achievements are enabled by a layer-based architecture (see Figure 3). Details of the different layers are given in the following subsections.

4.1 The Base Layer

The base layer of the SRA contains elements which are either CG or CV data. The architecture suggests that these elements are the smallest meaningful units that a capturing device can detect. We therefore name those units atomic scene elements, abbreviated ‘acels’. Each acel is coherent in itself, but independent from other acels. The number of dimensions an acel uses is conceptually unlimited. Possible dimensions are spatial and temporal dimensions, colours or reflectance. All common data types like images, meshes or videos are supported as acels, but any intermediate representation or additional dimensions on top of existing data types can easily be represented as well.

Many ideas for acel representations from Captured Video can be transferred from research on patches. Patches represent solid (sub-) surfaces for one animation/time instance of a scene. They evolve over time in a way which is plausible for human assumption, i.e. their position and shape are altered according to temporal and physical coherence. Patches represent physical entities and where introduced in the context of real-time reconstruction of
human faces, for example in [7]. Directly mapping the patch properties of acels shows that acels are well suited to represent solid and non-solid objects. They support physical coherences and the use of acels for numerous CG effects like relighting or shadows. Multiple more features can be easily added.

4.2 The Scene Layer

Multiple acels have to be registered in a global scene context. This registration is done in the scene layer of the SRA. The dimensions of a scene are the superset of all acel dimensions contained in a scene. Registration is not only done in space and time, but colour offsets and other measurement differences between acels can be corrected during registration. This component of the scene layer has therefore a structure comparable to a scene graph comprising multidimensional offset information in its branches. Sowizral proposed a method to present multidimensional scene volume information in graph structures [8]. In addition to placing acels in a global scene, the scene layer provides the lighting information for the scene. Lighting can be adjusted according to the scene lighting conditions independent of where acels were captured initially [9, 10]. Relations among acels are also expressed in the scene layer [11]. A coherency table expresses coherencies among the individual acel dimensions and features.

4.3 The Director’s Layer

The director’s layer defines the usage of scene content. The most important form of scene usage is scene perception. Cameras describe the traditional way of perception by defining intrinsic and extrinsic camera parameters. A novelty is that these virtual cameras are not limited to physical plausibility, but can feature several depth planes or shaped focal depth, introduce motion blurs, which are contradicting physical motion, or change the light sensitivity over one frame. Moreover, by defining user interaction rules, users further down the processing chain may be allowed to modify director’s decisions.

5 IMPLEMENTATION

In order to meet the requirements to the SRA defined by the intended advances and the algorithms presented above a flexible and extendable implementation is required. This becomes especially important as all components of the video processing chain are constantly enhanced and further developed. A possible implementation enabling these demands is an API to an underlying data structure. The SRA API is an implementation of the concepts presented in Section 4 making use of the content capturing methods presented in Section 3.

The underlying scene data can be any structured data that the SRA API supports. The structure represents the different layers of the SRA as well as scene elements such as acels, configuration data or interaction rules. The file readers which understand the underlying format are not part of the SRA API, but can be exchanged with the format of choice.

Support for a certain type of data is enabled in the SRA by the necessary converters in the API. These converters need to understand the data and be able to provide it to different processes in the API in the required representation. Exemplarily a mesh converter can be asked to return a mesh representation of an arbitrary input acel. If the data type of the input acel is supported as a mesh the request can be processed and an internal mesh representation can be provided. The set of converters can be arbitrarily extended to meet the data types of different data sources as well as the input requirements of further processing tools and applications.

Scene Modules in the API are used to initialize and execute computational processes. These modules can make use of converters and additionally implement further algorithms that process and enhance scene data. A module requires scene data in a certain format as input and can be triggered to be executed on demand. Exemplarily for such modules are getter-modules for textures or objects. These modules apply converters to transform acel information into a desired representation and present them to the next higher level as an object or a texture.

Figure 6: Structure of SRA Implementation
The third part of components contained in the SRA API contains interfaces for tools. Different applications have different demands to the Scene Representation. A certain tool interface can fulfill these demands by providing scene content application specific. Exemplarily a video rendering tool assures that accel data is presented frame based to a video renderer, which can then render the content of the scene per frame. Alternatively, a free-view interface can present the full content of a 3D scene and be rendered as a static scene to navigate in.

Neither the number of converters nor computational modules or application interfaces is limited. All of these can be extended with the growing demands from users, applications and algorithms. As such the implementation of an API for SRA access presents currently the perfect solution to have an extendable and flexible interface which does not limit the creativity of its users. The SRA API is a C++ library which can be included in applications in order to make use of the scene features. It can then be accessed by scripting languages (Python) or through the header functions exposed by the API. Thus it provides an easy to use interface to the scene developments.

6 VERIFICATION

A prototype to prove the conceptual ideas presented above was created. 100 frames of a billiard scene are represented in the SCENE layers and rendered. Figure 6 shows one of the video-clip frames, which exemplarily presents novel features and the paradigm change enabled by the SCENE format.

The prototype contains five acos: the static background, two independent players, the balls and the Minerva head. While the background is a single color bitmap plus depth, the players and balls have a temporal dimension as well. The Minerva head is a mesh with material properties to show the seamless integration of bitmaps and meshes in one single layout. It is relighted according to lighting conditions as described in Section 2. Knowing about the individual balls, camera effects like motion blurs can be applied to the pool balls.

The director’s layer describes a camera, which renders the scene off-angle to the original capturing device to make depth visible. Photorealistic camera effects can be synthetically added with the help of information captured as described in Section 2 [12, 13].

7 CONCLUSION AND FUTURE WORK

This paper presents a paradigm change in the way future video content can be produced to enable computational videography. Furthermore, a representation allowing this change from an architectural viewpoint is introduced. To our knowledge this is the first approach to redesign the full movie production process with the goal of enabling computational videography on multidimensional video content. Future work will need to further specify the SRA to meet the quality and algorithmic demands posed by content consumers and developers. Existing and novel ideas of computational videography can be designed to make use of the extra information provided through the SRA. Acquisition hardware will be designed to capture an ever increasing amount of multidimensional data for advanced video processing.

Some of this work is currently covered by SCENE project partners. Next to the five institutes mentioned in the list of authors the companies ARRI, Barcelona Media, Brainstorm and 3Dized are collaborating partners in the SCENE project. A full project description and the latest developments can be found online [1].

This work has been supported by the EC within the 7th framework programme under grant agreement no. FP7-IST-287639.

References