A Novel Representation for Digital Scenes

T. Herfet, C. Haccius*, V. Matvienko, Intel Visual Computing Institute, Saarbrücken, Germany
S. Fort, Barcelona Media, Barcelona, Spain

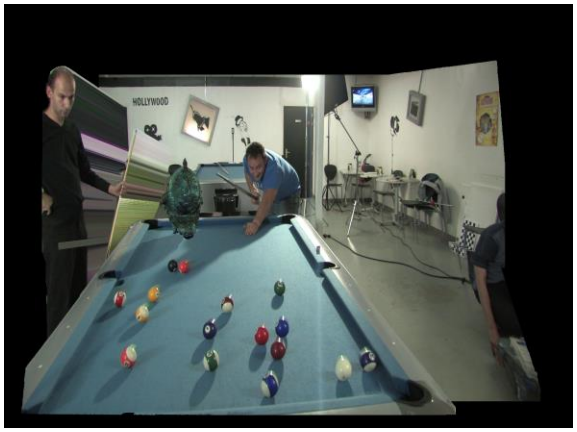

Figure 1: *Frame showing the different scene elements*

1 Introduction

SCENE is an ongoing research project dedicated to create and deliver richer media experiences [Hilton and Fuenmayor 2013]. A con-sortium of international research and industry partners aims to enhance the whole chain of multidimensional media production: new capturing devices, advanced processing tools and a dedicated renderer. Central is a novel scene representation which enhances and facilitates production processes of video content.

Current video production processes are still based on video acquisition methods that were shaped by physical bottlenecks of image capturing devices. The necessary exposure time of film or image sensor dictated shutter time or lens aperture. As a consequence loss of image quality due to blurs was in inevitable. However, image sensors and acquisition hardware have advancing and removed this physical bottleneck. As blurs have defined motion pictures they remain to be stylistic devices. Current video representations fail to make use of both: the advanced abilities of image capturing devices and the creative input of the cameraman.

We envision a scene representation that enables computational videography on captured video data and maintains the camera-man's creativity. Additionally, our representation meets several desires from movie processing and production. In detail these are:

1) A single format for all information necessary for video processing, such as camera calibration data, lighting information, spatial knowledge or coherency information.
2) Storage of captured and generated data in its best quality, so scaling or sampling of data is not required. Quality decreasing effects do not actually affect the captured data.
3) While traditional video content is frame based, we envision an object based representation. This facilitates object modification enables content interaction.
4) Captured video and computer generated data are currently living in very separate worlds. An early integration of both facilitates postproduction and makes computer generated content more realistic.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
SIGGRAPH 2013, July 21 – 25, 2013, Anaheim, California.
2013 Copyright held by the Owner/Author.
ACM 978-1-4503-2261-4/13/07

Figure 2: *Replacing segmented object and blur effect*

2 Our Approach

We have designed a layer-based scene representation architecture based on the idea of recreating virtual scenes for virtual movie production. We call the layers base-, scene- and director's layer.

The base layer corresponds to a collection of stage settings, stage props, actors and other objects. These elements are the atomic elements a scene is composited of; therefore named acels (**a**tomic s**c**ene **el**ements). Those acels are coherent in themselves, but independent of other scene elements. They can have any number of dimensions and are not limited in the information content.

The scene layer creates relations among the acels, thus placing the independent acels in a global scene context. The scene dimensions are a superset of all the acel dimensions occurring in a scene, and each acel is registered with respect to each dimension in the global context. Acel coherencies stored in this layer additionally assign relations to dimensions of different acels in a scene.

The third layer is the director's layer which contains the director's decisions. Most basic are camera parameters contained in calibration matrices. Furthermore, effects (like blurs, color offsets, etc.) are stored in this layer, modifying the underlying data representation but not the data itself. Last, interaction rules may allow user interaction with the scene elements.

A prove-of-concept implementation shows the effect of the architecture presented above. Figure 1 shows a billiard scene captured with color plus depth sensor. A computer generated mesh was relighted according to the scene lighting conditions and inserted above the pool table to demonstrate the merge of captured video and computer generated data. A small effect shows the use of the director's layer: The white ball is replaced by a black ball and a motion blur is added on top of the replaced object, as shown in Figure 2. This small implementation already shows that existing production steps are facilitated and new possibilities are created by the use of computational videography on this data.

References

HILTON, A., V. AND FUENMAYOR, E., 2013, Novel scene representations for richer networked media, http://3d-scene.eu, retrieved Feb. 2013.

*e-mail: haccius@intel-vci.uni-saarland.de