

Erfassung und Präsentation digitaler Replica von Artefakten des kulturellen Erbes

Acquisition and Presentation of Virtual Surrogates for Cultural Heritage Artefacts

Christopher Schwartz, Reinhard Klein
Universität Bonn
Friedrich-Ebert-Allee 144
53113 Bonn

Tel.: +49 228 73-{54183,4201}, Fax: +49 228 73-4212

E-Mail: {schwartz,rk}@cs.uni-bonn.de, Internet: <http://cg.cs.uni-bonn.de>

Zusammenfassung:

Das Erscheinungsbild eines Artefakts des kulturellen Erbes wird durch die Augen eines Betrachters wahrgenommen und kommt durch die Absorption, Brechung und Reflektion von Licht zustande. Dabei spielt nicht nur die drei-dimensionale Form sondern auch die verwendeten Materialien eine große Rolle. Um das digitalisierte Objekt einem Betrachter möglichst realitätsgetreu darstellen zu können sollte daher beides aufgenommen, gespeichert und auch wiedergegeben werden. Gebräuchliche Formen wie Fotos oder einfache (eventuell texturierte) geometrische 3D Modelle sind hierfür jedoch nicht ausreichend. Aus diesem Grund forschen wir an Geräten und Repräsentationen die eine realitätsgetreue digitale Darstellung, eine „digitale Replica“, erlauben. Die hier vorgestellte Technik ermöglicht die weitestgehend automatische Erfassung und Verarbeitung von kleinen bis mittelgroßen opaken Objekten, ein interaktives Betrachten der digitalisierten Objekte und die schnelle, progressive Übertragung im Internet.

Abstract:

The appearance of cultural heritage artefact is perceived by the eyes of an observer. The perception is caused by absorption, scattering and reflection of light. For this process, not only the three-dimensional form but also the materials the object consists of play an important role. To provide a faithful impression, both aspects, shape as well as appearance, have to be acquired, stored and rendered. However, conventional techniques, such as pictures or (possibly textured) geometric 3D models, are not sufficient for that task. Therefore, our research deals with devices and representations that allow for a faithful digital reproduction, a virtual surrogate. The presented technique facilitates the mostly automatic acquisition and processing of small to medium-sized opaque artefacts, the interactive inspection of the digitized objects and the fast and progressive transmission over the internet.

Introduction

The digitization of three dimensional objects is recently gaining importance in the field of cultural heritage (CH). There are many good reasons for this development. For instance, digitizing CH collections allows the safe and instant access to all items, be it for researchers, curators or the general public. 3D digitization may also serve a documentation purpose, capturing the state of an object prior and after restoration, before lending it to some other institution or for monitoring its decay. Objects digitized with high quality can be used as “virtual surrogates” for their physical counterpart. This does not only take risks from fragile or expensive objects, but it even allows for forms of dissemination that would be impos-



Figure 1: example of virtual surrogates acquired and presented with the proposed approach

sible with the physical object. The “virtual surrogate” can be showcased at arbitrary many locations simultaneously without any significant additional costs. It can easily be combined with other digital objects and put into different contexts, e.g. the historical use case or discovery on an excavation site. And last but not least, virtual surrogates can even be used for the public dissemination over the Internet, providing the capability to reach an enormous audience.

Especially in the field of cultural heritage, where even subtle details can drastically change the interpretation of an object, it is important not to rely on simplified or exaggerated representations but to capture and convey the full appearance. This is the visual impression a human observer has of the object. Note that the appearance of an object not only depends on the object itself but also on the viewing and lighting conditions under which the object is observed. The impression of an artefact lit by candles is different from its appearance under a cloudy sky, a showcase with neon bulbs or in bright sun light. Therefore, in order to fully understand and experience a three-dimensional object, the observer should be able to inspect it from all sides and to put it under different illumination conditions (see Figure 2). This is for instance comparable to holding the physical object itself in the hand, turning it around and holding it closer or further away or keeping the physical object fixed while moving a light source around the object. Even walking around an object that is standing in a showcase is a bit of the same experience, but definitely less immersive and satisfactory. In both cases, the inspection provides not only insight into the three-dimensional shape of the object but also in the materials it is composed of. Depending on the orientation towards the observer or the light source, wear and scratches are revealed or patina or dirt becomes apparent. For this, it is important to notice that the appearance of an object, i.e. its visual impression to an observer, comes from the interaction of light with the objects surface and interior. Light coming from all different directions is absorbed, scattered or reflected by the object and eventually, some of it reaches the eyes. For a faithful digital reproduction of the appearance, these interactions of light with the object’s surface have to be captured and later appropriately simulated during rendering.

The amount of light that is reflected towards the observer from the object’s surface is dependent on various variables: First, the position on the surface it was reflected from. Different points reflect light differently, for example because the surface is covered by a different material at that part of the object. Second, the direction of view as well as the direction of the incoming light is important. Specular materials, for example reflect much more light towards the observer when the directions are close to the specular reflection. Retroreflective materials, such as cat’s eyes or some minerals, reflect more light in case the light and view direction are close together. Even apparently dull materials become increasingly reflective when view- and light direction approach gazing angles. Third, light transport within the material also plays a role. The most obvious example would be glass, where most of the observable light actually comes from transmission through the object and very little from reflection on the surface. But even fur or fabrics transport light below the surface, emitting it from a different point than the one that was lit. Depending on the amount of light transport through the object, one would speak of transparent, translucent, sub-surface scattering or opaque materials. Finally, fluorescence, phosphorescence and changes in polarisation that can be caused by certain materials also affect the scattered light. To be truly general, all of the above effects would need to be captured and synthesized faithfully. Independently, measurements of subsets have been performed, in some cases requiring strict limitations e.g. with respect to the object geometry. However, a complete measurement as well as the necessary simulation reproducing the measured appearance is prohibitively costly, extremely time consuming and to the best of our knowledge an acquisition cannot be performed on arbitrary 3D objects with any one device, yet.



Figure 2: A virtual surrogate of a Ganesh figurine made from labradorite, a mineral showing a play of colours. This object exhibits drastic changes in appearance depending on the viewpoint and direction of illumination, changing from a greasy impression of soapstone to a bright blue gleam.

Instead, we concentrate on the practical capture and representation of objects consisting of a large subclass of materials that can usually be found in cultural heritage. We neglect fluorescence, phosphorescence and polarisation as well transparency or translucency. As a result, we are able to present a working pipeline to create faithful virtual surrogates of 3D objects made from opaque as well as locally sub-surface scattering materials (see Figure 1 for examples).

In the remainder of this paper, we will explain the individual steps and components involved in this pipeline and roughly follow the workflow that generates a virtual surrogate of a physical object. Furthermore, some results that have successfully been obtained will be presented and strength and weaknesses of the overall approach will be discussed.

A Matter of Representation

In order to faithfully represent the desired class of 3D objects, all necessary data for the simulation of the light interactions have to be either measured or derived. The types of light interactions can roughly be grouped by the size of the geometric structures they are influenced by. This concept is illustrated in Figure 3 and is for example described in [1]. First, there is the *macroscopic scale*. Features on this scale can be regarded to define the shape and geometry of a model. In the example in Figure 3, this would be the body, legs, arms, head or even nose and stylized hair of the depicted Buddha figurine. At this scale, features are commonly stored as polygonal meshes and acquired by a 3D scanner, e.g. laser-range scanners or structured light systems. The other extreme is the *microscopic scale*. On this level, the light is thought to interact with microscopically small facets on the surface of a material. These *microfacets* are orders of magnitudes below the resolution of the human eye. However, microfacets have a major influence on the view- and light-direction-dependent appearance of the surface: The amount of reflected light can be influenced by the orientations and the facets can occlude or cast shadows or inter-reflections on one another. Instead of acquiring, storing and simulating this incomprehensible amount of tiny structures directly, usually only the statistical distribution of the reflected light in dependence on the directions is modelled by analytical *Bidirectional Reflectance Distribution Functions* (BRDFs). As the material composition may vary over the surface of the object, different BRDFs or mixtures of them would have to be employed.

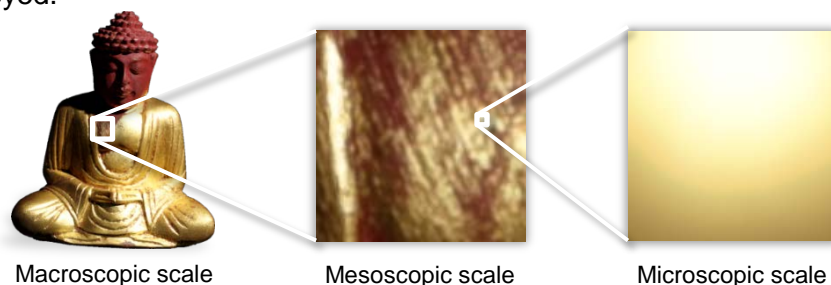


Figure 3: Illustration of the different scales considered for light interaction with an object

Inbetween the two extremes lies the *mesoscopic scale*. Features in this scale are just still visible with the naked eye but are usually not considered to be part of the defining shape of an object. The small bumps in the zoom-in on the surface in Figure 3, coming from scratches and brush strokes can be considered meso-scale features. Other examples might be small fibres in wool yarn or holes in the surface of an eggshell, which lead to its cavernous appearance. Capturing these fine structures with a 3D scanner and representing them directly with a polygonal mesh requires an enormous effort during scanning, a lot of storage-space for the representation and vast computational resources during rendering to really perform a fine-grained light simulation on that many polygons. However, using a distribution approach similar to BRDFs is also not feasible, as mesoscopic details can be clearly and individually identified by a human observer and, for example in the case of scratches, be very important to a CH specialist.

A reasonable solution to this problem is the use of a data-driven representation that directly stores measured data in a form that can be used for a faithful depiction without being computationally too demanding. A very common example would be the use of a texture, mapped onto the 3D geometry. The texture can be understood to be a 2D image with a certain resolution. Here, a single colour

value is stored per texture element (texel), i.e. pixel in the image. Mesoscopic features, e.g. scratches, can be measured and represented by taking photographic pictures and mapping them onto the correct position on the surface. This does not yield any considerable computational overhead and costs little more space than storing the images themselves.

However, a single colour value per texel is not enough. As argued before, the appearance of an object is dependent on the view-direction and its illumination. For this reason, we propose the use of the *Bidirectional Texturing Function* (BTF). A BTF can be regarded as a large collection of textures, which depicts the appearance of a surface under a set of different view and light directions. Similar to textures, the texels of a BTF are mapped onto the objects surface. Instead of a single colour value, the texels contain a function called *apparent BRDF* (ABRDF), which is described through interpolation of the discrete set of stored view- and light directions. The efficient acquisition, compression, transmission and rendering of BTFs have been the topic of many publications, making this technology a well-studied choice.

Another technique that facilitates the reproduction of appearance, which already has found application in CH, are the *reflectance transformation images* (RTIs), sometimes also referred to as polynomial texture maps (PTMs). Although, RTIs originally only considered different illumination directions, recently extensions to RTI have been proposed (e.g. [2]) that also take the view dependent effects into account by capturing multiple RTIs from different viewpoints. In this case, the capture process is essentially the same apart from the fact that no geometry is acquired. Nevertheless, for most applications, using a 3D geometry to describe the shape and a BTF for the small scale appearance presents a more balanced solution and yields considerable advantages regarding compression, rendering of novel viewpoints or illumination conditions (see Figure 5, right) and the composition of novel synthetic scenes.

Integrated Acquisition using a Multiview Dome Setup

A commonly used approach for the digitization of CH artefacts is to capture the geometry of the object and take additional pictures. The images need to be in alignment with the object so they can be used to reconstruct a texture. Methods, such as multiview stereo, can be used to reconstruct the 3D geometry directly from the images themselves, avoiding a separate alignment and hence facilitating an integrated acquisition of geometry and texture. The main problems with this strategy are that it often suffers from noise or holes in the 3D reconstruction and even more important that it fails to faithfully capture the appearance. This is mainly due to uncontrolled illumination conditions during the capturing phase and due to the sparse sampling of the appearance by a few images only. Therefore, such an approach is not really well-suited for high quality 3D digitization including appearance. To overcome at least the problems of insufficient geometry, off-the-shelf 3D scanners could be used for capturing the macroscale geometry with a high quality. However, aligning the pictures with the geometry is a non-trivial and tedious task. And just taking the tens of thousands pictures under controlled illumination conditions that are necessary for the reconstruction of a BTF cannot be performed manually.

Therefore, an integrated and automated approach that captures a high quality 3D geometry and the necessary pictures for reconstructing the BTF is desirable. Different designs for a setup that is capable to acquire pictures from view-directions all around the object, illuminate the object from various light-directions and support integrated 3D reconstructions have been proposed. They mainly vary in the degree of parallelism and hence acquisition times that can be achieved for a certain level of directional coverage. On the one side are setups that move a single camera and a single light-source around the object during acquisition [3], while on the other extreme a previous multiview/multilight setup of ours employs 151 consumer cameras [4]. Here, the cameras are mounted on a dome structure around the object, facilitating to take pictures from all view-directions simultaneously and using the camera strobes as a light source.

We propose a balanced solution that lies in-between the two extremes. Our multiview/multilight dome setup (depicted in Figure 4) employs eleven cameras that are mounted on an arc, providing some degree of parallelism, combined with a turntable for capturing the object from all sides. 198 LED light sources are mounted on the full dome, avoiding the need for mechanical movement for having light directions from all sides. To enable the integrated 3D reconstruction, four projectors that are used to project structured light patterns on the object are installed at different heights next

to the camera arc. The whole setup has a diameter of 2.30 m and a height of 1.50 m (without legs). It can be slid open for easy access to the sample-holder and the acquisition process is completely computer controlled and runs automatically without human supervision. The acquisition has to take place in a darkened environment, e.g. shielded by curtains, and is started remotely. Although the design choice of using only an arc of eleven instead of 151 cameras as our setup proposed in [4] reduces the amount of parallelism and slightly increases acquisition times, it has several advantages for practical use. First, it allows employing more expensive but also more stable and faster industry cameras and high quality lenses without costs shooting through the roof. Second, the stream of data coming in parallel from the eleven cameras during the measurement can still be handled by the single workstation computer that also controls the device, making the capture process more maintainable. Finally, having the cameras on a single arc allowed the dome to be designed lightweight and demountable into manageable pieces, rendering the proposed set-up movable and easy to build up.



Figure 4: the proposed multiview/multilight dome setup. Left: fully assembled, the front-left quarter is slid open to give a better view on the components. Right: the dome is disassembled and packed in a single transport van.

By design, integrated setups are very rigid and only acquire objects up to a certain size. For high quality results, the objects should also not be too small for the employed sensor-resolution of the cameras and focal length of the lenses. Using high quality 50mm prime lenses by Zeiss, and SVS Vistek cameras with 2048x2048 pixels resolution, our dome is suitable for objects with a maximum edge length of 24 cm, for example small figurines. On an object with this edge length, mesoscale structures of the size of 180 μm can still be resolved. By using lenses with different focal lengths, other sizes could be addressed as well. However, there is an upper limit due to the fixed distance of one metre of the cameras and light sources to the sample-holder at the centre of the dome.

Please note that some of the virtual surrogates depicted in Figure 1 were captured using the setup described in [4] and others using the proposed movable multiview/multilight dome.

From Images to a 3D Representation

After performing an average acquisition, more than 186,000 images of the object have been captured, occupying about 1.1 TB of disk space. Several consecutive processing steps have to be performed to boil down the massive amount of raw image data to a manageable representation consisting of a 3D polygonal mesh, serving as a proxy for the macro-scale appearance effects such as shape, shadowing and occlusion, and a matching BTF to depict the object appearance on the mesoscopic and microscopic scale. The details describing the processing of the data have been published in [4].

The first output is a polygon mesh, approximating the objects surface. As our intention is to only capture the macroscale geometry with the polygonal mesh, it is sufficient to store a comparably low number of polygons. From our experience, it is a good choice to store appearance details at the

size of 1 mm using the geometry. This results in meshes with about 100,000 triangles, occupying 10 MB of disk space, which can be directly used for real-time rendering.

The second result from processing the captured images is the BTF. To facilitate real-time rendering, which is crucial for the desired task of interactive object inspection, the appearance information conveyed in the images is re-sampled into a dense set of 151 view and light directions per BTF-textel. The directions are chosen to equidistantly sample the hemisphere above the digitized surface with a reasonable angular coverage. Without applying further compression, the BTF would thus be a stack of 22,801 view- and light-dependent textures: One for every possible combination of view and light directions. To fully exploit the available camera resolution, we usually create BTFs with 2048x2048 texels, which would require about 534GB.

Obviously, this amount of data is prohibitive for the economic storage of whole museum collections as virtual surrogates and way too much for loading even a single object onto the graphics card for real-time rendering. Therefore, the final BTF needs to be compressed. To obtain the necessary data-reduction a lossy compression is mandatory. For still remaining generic with respect to the objects materials we employ a data-driven approach called *factorization*. It is based on keeping only a certain number of c most significant components. This way, the amount of data can be drastically reduced and still contain the relevant appearance information. Conveniently, the appearance can be rendered directly from the compressed representation without too much computational effort. Interactive inspection becomes easily possible with 6 year old graphics hardware.

In our experience, $c=100$ components are usually enough to have no noticeable difference to the appearance in the photographic pictures themselves. A compressed file with this number of components is 1.6GB in size and can be used as a “master”-file to create lower-quality versions with less components for other applications, for example to support rendering the virtual surrogate on older hardware or mobile platforms.



Figure 5: the virtual surrogate consists of a polygon mesh and the BTF. The rendering of the virtual surrogate reproduces the appearance to a photographic picture from the measurement. An RTI created from the same measurement data fails to reproduce effects at the macroscale, e.g. the shadow of the head and hands, and at the microscale, e.g. the highlights of the red paint and gold.

Streaming of BTFs and Dissemination in the Browser

One possible field of application for virtual surrogates is the public dissemination. Especially the possibility of an interactive and immersive experience for the users over the internet is an interesting aspect of this technology. Despite the large file-size of the digital “master”-files for the virtual surrogates, we demonstrated in [5] that an effective and efficient progressive transmission of BTFs over the internet is in fact possible. The interactive exploration of virtual surrogates is enabled from directly within the website in the browser and can start after less than 1 MB of the BTF is transmitted, which takes only a few seconds with widely available DSL or 3G connections.

Figure 6 demonstrates the visual quality of the interactive exploration in the browser, in which the user is at all times able to freely change the viewpoint and light direction. The quality of the displayed visualization increases progressively during the interactive session, while more data is transmitted and integrated in the background. After about 7 MB are transmitted, the appearance reproduction is almost perfect and only the zoom-ins reveals that some fine details are still blurred with respect to the fully transmitted version.

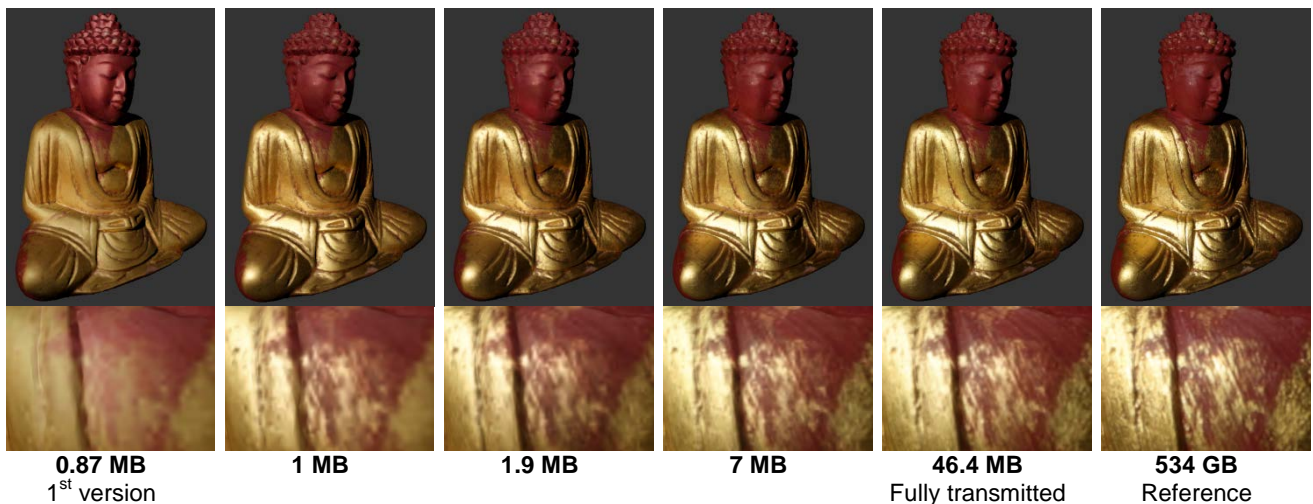


Figure 6: appearance quality increase of the virtual surrogate during progressive transmission over the internet. Please note that the images are screenshots of an interactively explorable 3D model. The lower images are zoom-ins to the chest of the Buddha figurine.

For using this interactive viewer, neither the host nor the client side need any kind of special equipment. On both sides, everyday web-technology, i.e. a standard HTML web-server and a fairly modern browser respectively, are the only requirements. For the progressive transmission the component data coming from the compressed “master”-file is converted into a set of PNG image-files and a JavaScript file describing the transmission order. For the example shown in Figure 6, $c=32$ components were used, resulting in 448 PNG-files with about 100KB each. At the client side, the browser simply requests the PNG files in the order given in the accompanying JavaScript file. By utilizing WebGL, a technology supported by most current browsers even on mobile platforms, the content in the PNGs are decoded in the background and the BTF is rendered in real-time using the graphics hardware.

The Pipeline from a User Perspective

Performing an acquisition does not require any expert knowledge but a short introduction to the control software. Just a few basic parameters have to be chosen by the operator prior to the measurement. For instance, many materials exhibit a dynamic range that exceeds the capabilities of the cameras, i.e. they reflect very little light for some directions and produce bright highlights in others. To cope with that, multiple images with different exposure settings have to be acquired. Furthermore, some objects exhibit a more complex shape or a more complex angular dependent appearance than others. In these cases, the number of turntable rotation steps can be increased to acquire the object from more angles. Therefore, the operator has to choose the number of exposures and rotation-steps of the turntable. The duration of the measurement and the amount of generated raw data directly depends on these values as well. An average acquisition generates more than 186,000 images of the object in about 7 hours. The raw measurement data takes up to 0.9 TB of storage capacity and is hence directly captured onto an external hard disk. This disk can then be connected to another computer to process the data into its final form, a polygon mesh and a compressed BTF, requiring only 1.7 GB. These files can directly be used for storage and high quality reproduction or for the generation of derived versions, for example a 46 MB PNG file representation for the transmission over the Internet.

During the post-processing the dome device itself can be used with another disk to perform the next acquisition in parallel. On a high-end workstation PC (2 hexacore Intel Xeon 2.4GHz, 144 GB RAM, GeForce GTX 570 GPU) the post-processing takes about 36 hours. Although every processing step can be controlled by a number of additional parameters, the employed algorithms are reasonably robust to use a conservative set of parameters for a wide range of objects. This allows the processing to take place automatically but still enables the operator to intervene and improve the results, if necessary.

To bring the movable dome to a different location, e.g. for on-site acquisition, it can be taken apart into four gantry pieces and six transport boxes with equipment. In this configuration all parts fit through standard doors, elevators and staircases and can be packed in a single transport van (see Figure 4, right). Disassembly and assembly can be performed by two trained technicians, taking one and two days respectively.

After assembly the device needs to be calibrated. This process is automated as well and simply requires presenting two special calibration targets to the device. Including computations, this process takes about 4 hours. In case the device needs to be re-calibrated on-site, this can be performed by the same personnel that are trained in the control software. However, if the device is handled with the necessary care, the calibration is quite stable over a rather long period.

Conclusion

We presented a feasible and practical pipeline for the acquisition and presentation of virtual surrogates, exhibiting a faithful appearance under novel viewing and lighting conditions. Although far from being completely general, the presented approach covers a large class of objects commonly encountered in cultural heritage. While it is not applicable for objects that are large in size or are made from transmissive materials, it works well for small to medium-sized artefacts with an edge-length up to 24 cm (or more depending on the employed lenses) that are opaque or locally sub-surface scattering. Objects can be acquired and processed automatically, requiring very little human oversight. Although the average acquisition of 8 hours takes longer than just using a laser scanner, the additional value gained by being able to faithfully reproduce the overall appearance certainly outweighs this aspect. The proposed acquisition setup is easy to operate and can be taken into parts and assembled on-site rather than moving precious cultural heritage items to the location of the setup.

Instead of following a purely image-based approach such as multi-view RTIs, we make the deliberate choice of having a 3D polygon mesh representing the macroscopic aspect of the appearance and the image-based BTF for the mesoscopic and microscopic appearance effects. By supporting a progressive streaming approach to explore the digitized artefacts directly in the browser, the proposed representation is well suited for massive dissemination. While keeping the captured raw data would be prohibitively costly, a processed virtual surrogate is capable of closely reproducing the appearance of the captured pictures and at the same time yields an effective compression ratio that allows the storage of almost 2,000 objects on a single hard-disk. The need to compromise between accuracy and manageable effort can in the end of course only allow for an approximation of the mesoscale and microscale appearance. However, in our experience the chosen set of captured directions and the lossy compression work reasonably well for the faithful reproduction of a large number of materials.

In conclusion, the presented approach can be considered as a viable choice for the high quality digitization for many artefacts encountered in the collections of cultural heritage institutions. Moreover, some examples depicted in Figure 1 show that the proposed technology can find application outside the cultural heritage sector as well.

Bibliography

- [1] S. H. Westin, J. Arvo und K. E. Torrance, „Predicting reflectance functions from complex surfaces,“ in *SIGGRAPH*, 1992.
- [2] P. Gunawardane, O. Wang, S. Scher, I. Rickards, J. Davis und T. Malzbender, „Optimized Image Sampling for View and Light Interpolation,“ in *Proceedings of VAST 2009*, 2009.
- [3] M. Holroyd, J. Lawrence und T. Zickler, „A Coaxial Optical Scanner for Synchronous Acquisition of 3D Geometry and Surface Reflectance,“ *ACM TOG*, Bd. 29, Nr. 4, pp. 99:1-99:12, July 2010.
- [4] C. Schwartz, M. Weinmann, R. Ruiters und R. Klein, „Integrated High-Quality Acquisition of Geometry and Appearance for Cultural Heritage,“ in *Proceedings of VAST 2011*, Prato, Italy, 2011.
- [5] C. Schwartz, R. Ruiters, M. Weinmann und R. Klein, „WebGL-based Streaming and Presentation Framework for Bidirectional Texture Functions,“ in *Proceedings of VAST 2011*, Prato, Italy, 2011.