DIGITIZATION OF PEOPLE AND OBJECTS FOR VIRTUAL MUSEUM APPLICATIONS

Ingo Feldmann[a], Oliver Schreer[a], Thomas Ebner[a], Peter Eisert[a], Anna Hilsmann[a], Nico Nonne, Sven Haeberlein[b]

[a] *Vision and Imaging Technologies Department, Fraunhofer Heinrich Hertz Institut, Germany, ingo.feldmann@hhi.fraunhofer;* [b] *Trotzkind GmbH, Germany, nico@trotzkind.com*

Abstract: We present a system for the digitization of real people and objects for the integration into computer-generated virtual scenes. We target the creation of natural and realistic representations of historical sites, artifacts and objects, which can be integrated into virtual museum applications. Using Virtual Reality (VR) glasses, a user can stroll through a virtual exhibition and get an immersive impression of the artifacts. Our system allows to capture and digitize static as well as moving people and objects. In this way additionally a historical context can be generated in the virtual scene which fits to the artifacts and objects. Actors can be inserted to reconstructed historical sites to create a realistic and convincing historical experience. Virtual guides could provide additional information about the exhibits and enrich the scene. From a technical point, our system combines computer graphics and image-based rendering tools to represent real persons through realistic and natural moving 3D models. In addition, methods for the passive digitization of highly detailed 3D models have been developed.

1. INTRODUCION

Virtual Reality (VR) applications allow to generate a completely novel user experience. Based on computer generated graphical contentent an arbitrary scenario can be modeled and animated. New VR glasses technologies allow the user to experience this content in an immersive and direct way.

However, one main restriction of conventional VR compositing tools is their limitation to computer generated content. In many cases a time concuming modelling and animation process needs to be performed to create related VR scenes.

In contrast to this, in this paper we will propose a method which captures real people and objects and automatically converts them to 3D models. These models can be directly inserted to VR scenes.

This new way of digitizing real world content will be discussed in the context of virtual museum applications. Our vision is that by using VR glasses the user can virtually walk through a virtual exhibition and get an immersive impression of the artifacts.

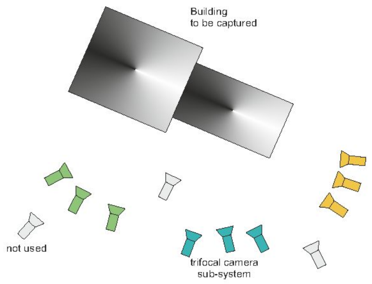

***Fig. 1:** Capturing of static objects with single camera from multiple perspectives*

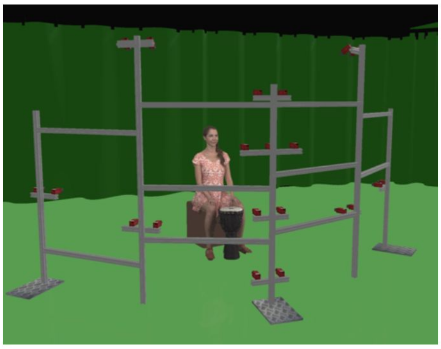

***Fig. 2:** Capturing of real persons with multi-camera system*

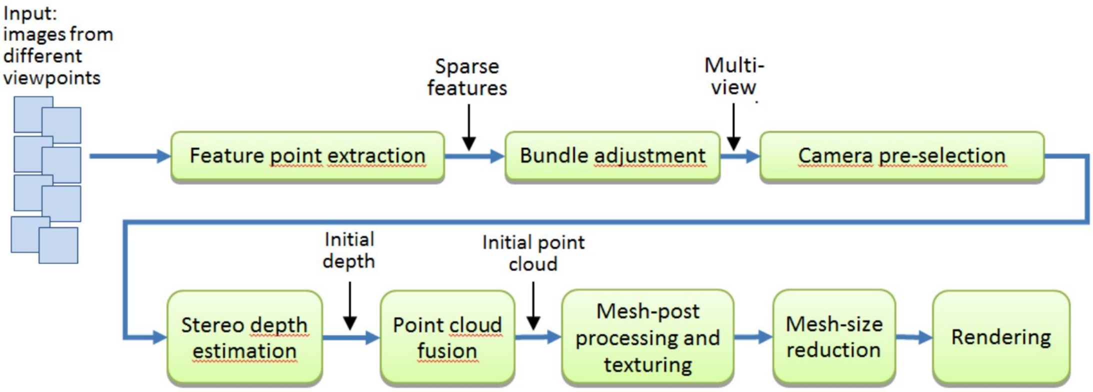

Fig. 3: Image based 3D data acquisition workflow

In this way additionally a historical context can be generated in the virtual scene which fits to the artifacts and objects. Actors can be inserted to reconstructed historical sites to create a realistic and convincing historical experience. Virtual guides could provide additional information about the exhibits and enrich the scene.

In the following, we will describe the technical background and basic functionality of the capturing and 3D reconstruction process for static 3D objects as well as real persons. The general algorithmic workflow is in both cases very similar whereas the data aqusistion itself differs. For static objects a single camera is sufficent. As illustrated in **Fig. 1** the object of interest needs to be captured from multiple positions in order to reconstruct the 3D geometry.

In contrast, for moving objects, such as persons, a fixed multi-camera setup as shown in **Fig. 2** is required. As shown in the figure, mulitple video cameras are arranged circularly around the person of interest. It is captured simultaneously from multiple angles and positions.

In the next section, first the general algorithmic workflow will be described for the 3D acquisition of moving objects such as people. The static acquisition case is discussed afterwards in section 3. Finally, potential use cases and results will be discussed in section 4.

2. MULTI-VIEW 3D RECONSTRUCTION

From a technical point of view, our proposed system is based on a multi-camera capture setup as shown in **FIG.** 2. As illustrated in the figure several camera pairs are grouped around the person of interest. Each camera pair serves as a stereo camera base unit. Each base unit allows to reconstruct the related scene structure information applying an image based depth estimation process. In a subsequent step the information of all stereo base units will be fused to one common 3D model.

Fig.3 shows the complete technical workflow for an image based 3D data acquisition. In order to estimate and fuse depth information an initial 3D camera calibration needs to be performed. It consists of a feature point extraction and a subsequent bundle adjustment step. An optional camera preselection step allows to reduce computational complexity by selecting only relevant cameras.

In the illustrated workflow chain we have used the methods described in 1 to gain an initial depth map of the actor. Based on the methods described in 2 we fuse the initial depth maps to a common overall 3D model. The resulting 3D mesh will be post-processed in order to remove artifacts and outliers.

Finally, the geometrical complexity of the 3D mesh model needs to be reduced in order to prepare the data for rendering in VR glasses applications. Here, initially a screened Poisson surface reconstruction is applied, which already significantly reduces mesh complexity. In addition, this step generates a watertight mesh. Holes that remained in the surface after the reconstruction are closed 11. Subsequently, the triangulated surface is simplified even further to a dedicated amount of triangles by iterative contraction of edges based on quadric error metrics 12. For restoration of details, which got lost during simplification, the utilisation of a texture in contrast to the vertex colours in the fusion result is required. The texture generation phase includes mainly two steps: First, the parameterization step, in which UV coordinates are calculated. And second, the texture creation step, in which the texture file is created and filled with colour values. This can be achieved

by either sampling the vertex colors of the initial 3D model or projecting and merging the captured images onto the simplified mesh.
Please note that a more detailed technical description of our used camera based 3D reconstruction of people can be found in 3.

3. STATIC OBJECT MODELLING

For the creation of static 3D scene models, we capture the artifacts with a DSLR camera from different viewing directions and use a warp-based approach 7 for 3D reconstruction, that exploites the entire image information for highly detailed models. The approch follows a three step approach:

1. *Initialization:* a rough object model is constructed from sparse point correspondences.
2. *Depth refinement:* surface geometry is refined by matching dense stereo pairs
3. *Model fusion:* several stereo reconstrcutions are fused to a full 3d model

These steps are described in more detail in the following:

3.1 GEOMETRY INITIALIZATION

Given a set of images captured from the object to be reconstructed, first, sparse point correspondences between feature points in the images are established. We use SIFT features 4 combined with a novel spatially aware matching technique 6 that creates a larger number of feature matches with higher reliability compared to standard matching approaches. From the image correspondences, camera positions and a sparse 3D point cloud are estimated using standard bundle adjustment while the camera intrinsics are determined via model-based camera calibration 5. The point cloud is then triangulated and a rough surface is fitted through the sparse positions similar to 8.

3.2 DEPTH REFINEMENT

Starting with the rough surface description from the point cloud, the geometry is refined using pairwise image warping. For a particular reference view I, the vertices of the surface mesh are optimized along the projection rays (depth direction), such that the image I, warped into the viewing direction of a second view J, matches the captured original view J as close as possible.

$$I(x) = J(W(x, d))$$

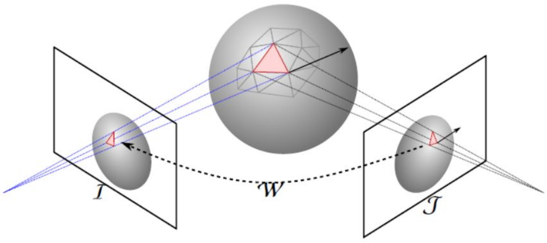

Fig. 4: The warp function W describes the mapping of the second image J onto the first one, I. It is parametrized by the depths of the mesh vertices along their projection rays.

Thus, the dense warping W is a function of the unknown vertex distances in the reference view I and is optimized using an optical flow-based optimization scheme 7. Using the optical flow constraint together with the mesh-based warping function, a highly overdetermined linear system of equations can be setup that is efficiently solved with a sparse solver. The large number of equations each corresponding to an image pixel lead to robust solutions while the use of image-gradients in the matching function allows for sub-pixel accurate image matching.

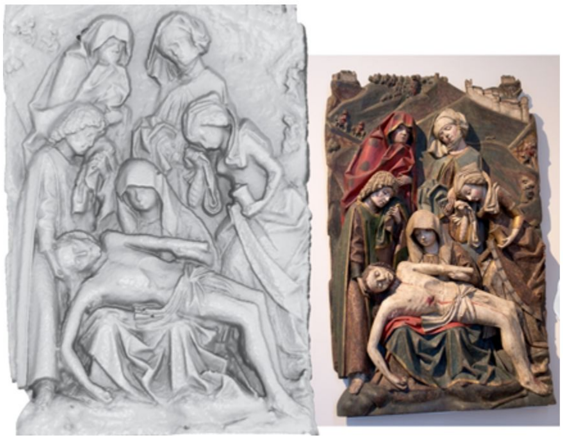

Fig. 5: Reconstruction from two frontal views.

Although the image matching is constrained by a piecewise affine motion along the epipolar geometry, untextured image areas can lead to mismatches. Therefore, smoothness priors are added to the cost function to be optimized. In contrast to global smoothness terms like the uniform Laplacian, we use a non-linear trilateral constraint motivated by the bilateral filter 9, which uses three different kernels for computing smoothing weights: weighting with the distance of a surface point from the filtered vertex, color similarity to allow for sharp discontinuities at color changes, and depth similarity preventing the regularizer from smoothing over strong depth gradients. With this additional term, fine details can be reconstructed while reducing outliers in homogeneous regions as illustrated in **Fig. 5**.

3.3 MODEL FUSION

The depth refinement method described above is able to compute depth estimates independently for each reference view. In order to create a complete 3D model, the different estimates need to be fused to a consistent description. For that purpose, the consistent mesh topology from the initialization is used, the individual depth estimates from the reconstructions are projected onto this surface, and then the best vertex candidate is selected from the existing samples as illustrated below (**Fig. 6**).

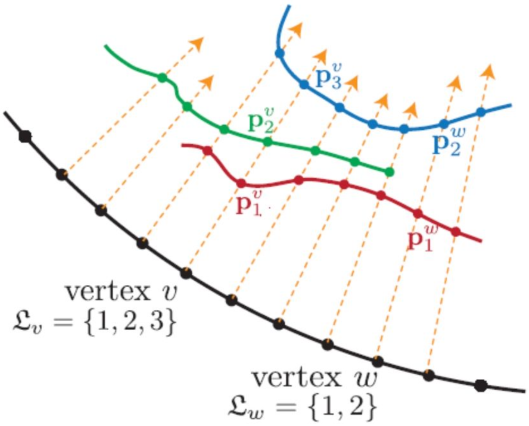

Fig. 6: *Illustration of the fusion of three contradicting surfaces to a consistent model.*

For the selection of an optimal vertex position, a loopy belief propagation 10 is used that optimized unary and pairwise cost terms for the vertices. These include a quality value for each depth estimate determined during depth estimation, geometric smoothness between vertex pairs and texture consistency between the reprojected views. Discreet optimization leads to a set of mesh labels that is used to fuse the individual meshes into a consistent 3D model (**Fig. 7**).

The result for a fully reconstructed static 3D object can be seen in **Fig. 8**.

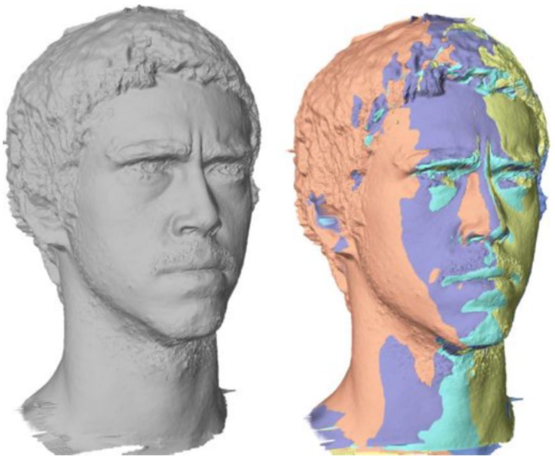

Fig. 7: *Left: Mesh fusion result from four source meshes. Right: Color coding of source mesh index.*

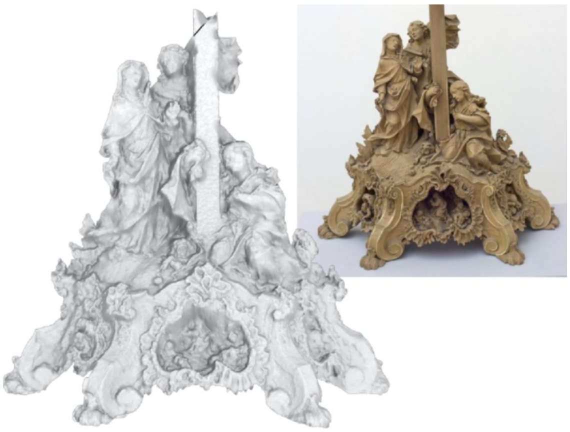

Fig. 8: *Example for highly detailed 3D digitization of static models*

4. USE CASES

Within the scope of museums and cultural heritage, there is a wide range of use cases for reconstructed people and objects.

All museum objects have to be placed in a context, which could be their origin (geographical context), period of production (chronological context), or their usage (qualitative context). Since many objects are unique, they cannot be easily used and placed in multiple contexts.

Using reconstructed objects in Virtual or Augmented Reality, offers the ability to create an arbitrary number of environments with different or even changing contexts.

If, for instance, an environment for the rosetta stone should be created, the stone could be shown in its geographical context, in this case it could be a place in Memphis, Egypt, since the stone is inscribed with a decree that was issued there. Since the decree came from King Ptolemy V in 196 BC, the viewer could be interested in understanding the time, people and surroundings better. By recreating a place based on still (partially) existing buildings and adding people to the scene (filmed with the method described before, re-enacted by actors dressing, behaving and speaking according to current research) the viewers could see its origin and gain a better understanding of the time and its customs.

But there are more contexts in which the rosetta stone can be placed. The stone was rediscovered in 1798 during Napoleons's campaign in Egypt in Fort Julien, near the city Rosetta. It is just as plausible, that the viewer wants to explore that context. Wanting to see the fort and the discovery itself, or how Napoleon inspected the stone himself. Of course there is also the more obvious context, in which the viewer could be interested in learning what makes that stone so famous today.

Creating those different environments virtually adds another benefit: instead of chosing for the viewers which environment they get to see, we can let them choose. The experience can be completely interactive and they can choose whether they want to learn something about Ptolemy, Napoleon, Ancient Egyptian hieroglyphs, Demotic script, Ancient Greek, or why it was so important to have them on one rock stele.

Since the viewers now have the ability to choose, we can also add additional informations, if a viewer is especially interested. Data that might already exist in a museum information catalog, but is currently not shown in the exhibition itself, can now be presented. Similarly tags and identifiers from such a catalog can be used to show other objects that are linked to the currently viewed object.

Other benefits are that virtual objects will always be there for the viewer – independent of restorations, traveling exhibitions, research, or even because part of the museum is under construction.

There is also no space limit. All objects from the museum depot can be shown. And they can be shown close-up. No glass, no barriers, no boundaries. In **Fig. 9**, an example for a Virtual Museum is shown that has been created with Unity Render engine and reconstructed objects. Actually, they cannot only be shown, they could also be picked up, viewed from any angle and even be used, as long as that interaction has been included.

Another factor to consider is the storytelling. While audioguides have found their place in the modern museum landscape, it is still different if a person standing next to the viewers is telling them about an object on display – especially if that person has a personal connection to the object, like having been part of the excavation or restoration.

In **Fig. 10** and **Fig. 11**, virtual scenes are presented that show the inclusion of reconstructed 3D objects and reconstructed moving persons in historical scenes.

This guide (who has been filmed and 3D reconstructed, as previously discussed) can even be made interactive and talk about subjects that the viewer is most interested in, similar to an expert from a guided tour – every viewer gets their personal guide.

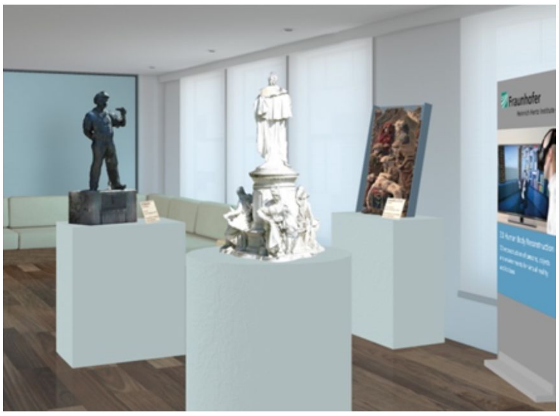

Fig. 9: *Example for Virtual Museum with several reconstructed objects*

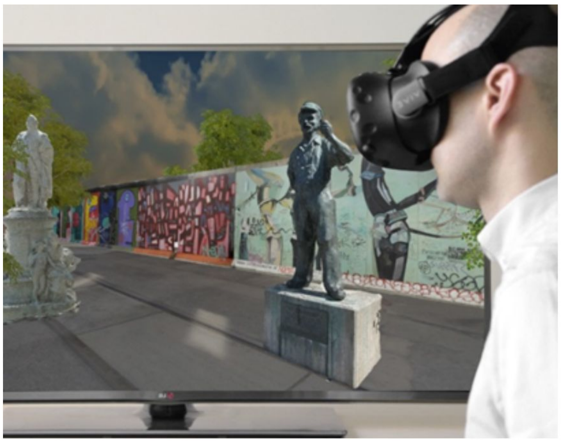

Fig. 10: *Example for reconstructed statues in historical context in a VR application*

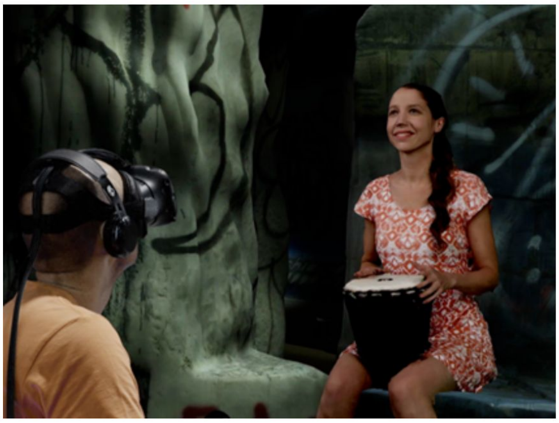

Fig. 11: *Example for reconstructed and enriched historical sites in Virtual Reality: A drumming actress was inserted next to the historical 'Berlin Wall'*

5. CONCLUSION

In this paper, a complete workflow for the creation of an immersive Virtual Museum experience has been presented. By using advanced Computer Vision and Computer Graphics technologies, it is possible to create arbitrary virtual scenes that include photo-realistic models from real cutural heritage artifacts and dynamic 3D models of persons.

Our proposed basic 3D reconstruction approach is similar for static objects as well as dynamically moving persons. The main advantage is that all the processing is performed automatically without any specific user interaction.

The presented reconstruction approach offers a variety of novel use cases and scenarios for museums, education and edutainement. Cultural heritage artifacts can be presented that never have been shown to the public before. Historical sites can be created and enriched by still existing artifacts and virtual guides can tell stories and explain the scene as if the viewer would be present at the place.

6. ACKNOWLEDGEMENT

Financial support by the European FP7 project BRIDGET (No.610691) and the H2020 project REPLICATE (No. 687757) is gratefully acknowledged.

7. REFERENCES

1. W. Waizenegger, I. Feldmann, and O. Schreer, (2011), "*Realtime patch sweeping for high-quality depth estimation in 3D videoconferencing applications*," in Proc. of Real-Time Image and Video Processing, San Francisco, California, United States, 2011.
2. S. Ebel, W. Waizenegger, M. Reinhardt, O. Schreer, I. Feldmann, "*Visibility-driven patch group generation*", Proc. of Int. Conf. on 3D Imaging (IC3D), November 2014, Liege, Belgium.
3. W. Waizenegger, I. Feldmann, and O. Schreer, "*Real-time 3D body reconstruction for immersive TV*", Int. Conf. on Image Processing, Phoenix, AZ, USA, Sept. 2016.
4. D. Lowe, *Object recognition from local scale-invariant features*, *International Conference on Computer Vision,* Corfu, Greece (September 1999), pp. 1150-1157.
5. P. Eisert, *Model-based Camera Calibration Using Analysis by Synthesis Techniques*, Proc. Int. Workshop on Vision, Modeling, and Visualization (VMV), pp. 307-314, Nov. 2002.
6. J. Furch, P. Eisert, *An Iterative Method for Improving Feature Matches*, Proc. Int. Conf. on 3D Vision (3DV), Seattle, USA, June 2013.
7. D. Blumenthal-Barby, P. Eisert, *High-Resolution Depth For Binocular Image-Based Modelling*, Computers & Graphics, vol. 39, pp. 89-100, Apr. 2014.
8. M. Kazhdan, M. Bolitho, H. Hoppe. *Poisson Surface Reconstruction*, *Symposium on Geometry Processing 2006*, 61-70.
9. C. Tomasi, R. Manduchi, *Bilateral Filtering for Gray and Color Images*, Proc. ICCV 1998.
10. S. Nowozin, C. Lampert. *Structured prediction and learning in computer vision*. Foundations and Trends in Computer Graphics and Vision, 6(3-4):3–4, 2011.
11. M. Kazhdan, and H. Hoppe, "*Screened Poisson Surface Reconstruction*", ACM Transactions on Graphics (TOG), Volume 32, Issue 3, June 2013.
12. M. Callieri, P. Cignoni, F. Ganovelli, C. Montani, P. Pingi, R. Scopigno, "*VCLab's tools for 3D range data processing*", VAST'03 Proceedings of the 4th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage, Brighton, UK, 2003.