Mixed Reality for Intuitive Photo-Realistic 3D-Model Generation

Wolfgang Sepp, Tim Bodenmüller, Michael Suppa, and Gerd Hirzinger

Institute of Robotics and Mechatronics
German Aerospace Center (DLR)
Münchner Str. 20, 82234 Wessling, Germany Tel.: +49 (0)8153 / 28 3473
Fax: +49 (0)8153 / 28 1134
E-Mail: wolfgang.sepp@dlr.de

Abstract: Appropriate user-interfaces are mandatory for an intuitive and efficient digitisation of real objects with a hand-held scanning device. We discuss two central aspects involved in this process, i.e., view-planning and navigation. We claim that the streaming generation of the 3D model and the immediate visualisation best support the user in the view-planning task. In addition, we promote the mixed visualisation of the real and the virtual object for a good navigation support. The paper outlines the components of the used scanning device and the processing pipeline, consisting of synchronisation, calibration, streaming surface generation, and texture mapping.

Keywords: VR/AR, 3D modelling

1 Introduction

In this paper, we focus on hand-guided digitisation of real objects. In contrast to the scanning on pre-determined trajectories, this approach is much better suited to objects with complex surface shapes because the vantage point can be freely chosen. Moreover, hand-guided digitisation allows to scan structural details with a higher resolution than other object parts, e.g., by adapting the distance to the object.

In this process, an immediate feedback of the quality of digitisationis very important. The quality of the resulting 3D model depends on the complete tool-chain, i.e., depth sensing, surface reconstruction, and texture mapping. If the latter two steps occur not on-site, then a tedious and time consuming re-scan or continuation of the scanning process is required. Therefore, we develop a processing pipeline capable of the concurrent 3D reconstruction and appropriate visual feedback.

In recent approaches and commercial applications, the problem is handled in different ways. For the digitisation of mid-scale (rooms) and large scale (buildings, cities) objects all scan paths are planned before digitisation, e.g., viewpoints or flight trajectories, and the 3D surface reconstruction occurs in a post-processing step after the thoughtful acquisition of depth information, e.g., [SA00], [AST+03], and [HBH+05]. Here, the data is usually gathered by large-distance laser scanners, e.g., the ZF Imager[1] or airborne camera systems.Any missing information (views) in the scene, which is discovered in the post-processing step, requires a cost intensive return to the site with the equipment.

[1] Zoller+Fröhlich Imager 5006, http://www.zf-laser.com, 2009

On a small scale, numerous commercial scanning devices exist. Usually, laser-stripe devices are used and also here, 3D surface reconstruction is done in a post-processing step, e.g., the Metris Modelmaker[2], the Faro Laser Line Probe[3], and [FFG+96]. The feedback to the person moving the scanning device is a 3D view of the captured data points on a display. The person scanning the surface has to infer more or less precisely from the shown 3D point cloud to which real surface points the shown data refers to. Additional help is given by reference points projected on the object requiring the operator to keep the scanning device in an appropriate distance to the object. A significant improvement is obtained by showing a partial 3D surface reconstruction of the most recently scanned surface patch, e.g., [HI00].

Here, we present the use of a continuous, streaming 3D reconstruction *c*oncurrently to the hand-guided scanning process. Accordingly, the user gets immediate feedback of the surface reconstruction result and can adjust his motion to scan objects parts that have not yet been digitised. Moreover, we gather live images from a colour camera in the scanner device, show them on the computer display together with a rendered image of the 3D model generated so far. This mixed reality extension allows the user to intuitively scan, evaluate, and continue scanning the objects on the desired locations.

In the following Section 2 the process of hand-guided scanning is analysed in depth and requirements for scanner systems are derived. In Section 3, the used hand-held device is introduced and aspects of synchronisation and calibration are discussed. Further in Section 4 and 5 the used methods for surface reconstruction and texture mapping are summarised. Section 6 reports technical aspects of the integration and presents modalities in a mixed-reality user interface that allow for an intuitive scanning process.

2 On Man-Machine Interfaces for Intuitive Hand-guided 3D Scanning

Digitising complex objects with surface scanner systems is a tedious process since multiple measurements from different views are needed in order to scan the entire object. The operator of a hand-guided system needs to identify suitable viewpoints and has to navigate the device to each of these views. Therefore, he has to keep track of the object parts that are already digitised and of those parts that remain to be scanned.

We identify two aspects in this task that are performed alternately by the operator: view planning and navigation. The view planning comprises the inspection of the already digitised parts and the determination of the next view. The navigation is the movement of the scanner around the object towards the view. Usually, this results into continuous movements along the surface.

In the following, we discuss possible modalities that support the scanning process in the mentioned aspects. With these insights, we derive system requirement that best suit the task.

[2]Metris Modelmaker, http://www.metris.com, 2009

[3]Faro Laser Line Probe, http://www.faro.com, 2009

2.1 View Planning

In order to plan a new scanning view (or sweep) over the object, it is mandatory to know, which parts of the object still have to be digitised. When scanning objects for the first time, it is impossible for the system to know the missing parts before scanning them. Hence, the system can only enumerate the complementary set of surface parts, i.e., those parts that have been successfully digitised.

This set does not necessarily correspond to the object parts that were located in the individual field of views. Local surface reflectance, for instance, constitute a natural limitation on ability to measure distances from specific views. Other views on the same surface patches might be necessary for a successful digitisation. Accordingly, it is more appropriate to record the scanned surface areas then to enumerate the past viewing positions.

In general, there are two possibilities to plan the next best view (or sweep) for the scanning process. Either the task could be done autonomously by the system or by the operator himself. The former task is known to be NP-complete. We expect the operator to perform better than an uninformed, autonomous exploration algorithm, because humans quickly perceive a rough 3D model from short views of the object.

Accordingly, we expect the operator to plan appropriate views on the object step by step based on the cumulative results from the past scanning actions. The cumulative information needs to be communicated to user appropriately, e.g., by showing a partial 3D model of the object.

2.2 Navigation

Moving the scanning device on a planned trajectory typically requires the operator to localise way-points of his plan on the real surface. The operator has to establish the link between the real scene and the reference of his plan, i.e., a digital representation of the scene. Hence, he has to link the virtual world to the real world. More precisely, three entities of the scanning setup have to be spatially referenced to each other: the object, the scanning device and the operator.

The user interface can support the navigation task by communicating the spatial relationship of these three entities. The simplest implementation, however, visualises only the scanned parts of the 3D object, leaving it to the user to locate himself and the scanning device with respect to the virtual object. A more advanced system renders the scene from the viewing position of the scanning device. In the ideal case, all entities are rendered from the current viewing position of the operator.

2.3 Requirements

System requirements are identified in the spatial, temporal, and radiometric domain. Essential parts of the system are a depth measurement unit and a spatial referencing unit that relates each depth measurement to each other. Photometric measurements are needed in order to capture the real textures of the object.

Apart from the sensor components, a consistent 3D model is obtained only when two spatio-temporal constraints are met. Firstly, each unit needs to be well calibrated. Secondly, the depth measurement and the spatial referencing information have to be timely coherent. These constraints are set up for photometric measurements and the spatial pose information, respectively. The hardware needs to support synchronisation and the streaming transmission of real-time sensor-information. The exigencies on synchroneity, however, depend on the velocity of the movements of the depth sensing and photometric sensing units.

3 The Multi-sensory 3D-Modeler

The DLR 3D-Modeler [SKL+07] is a multi-sensory scanning-device, which has been designed for the generation of photo-realistic models of small and mid-scale objects. The device integrates 3 types of range-sensors and allows to gather range information while moving the device. In order to consistently merge the information from different sensor units, a spatial calibration of the range-sensors with the external pose measurement unit is required as well as the time synchronisation of all units in the device. Special attention is given to calibration and synchronisation because they determine the final accuracy of the 3D reconstruction.

3.1 Components

The hand-held 3D-Modeler shown in Figure 1 consists of a rotating laser-range-scanner, two colour cameras, and two laser-line modules. These components allow for three types of depth-

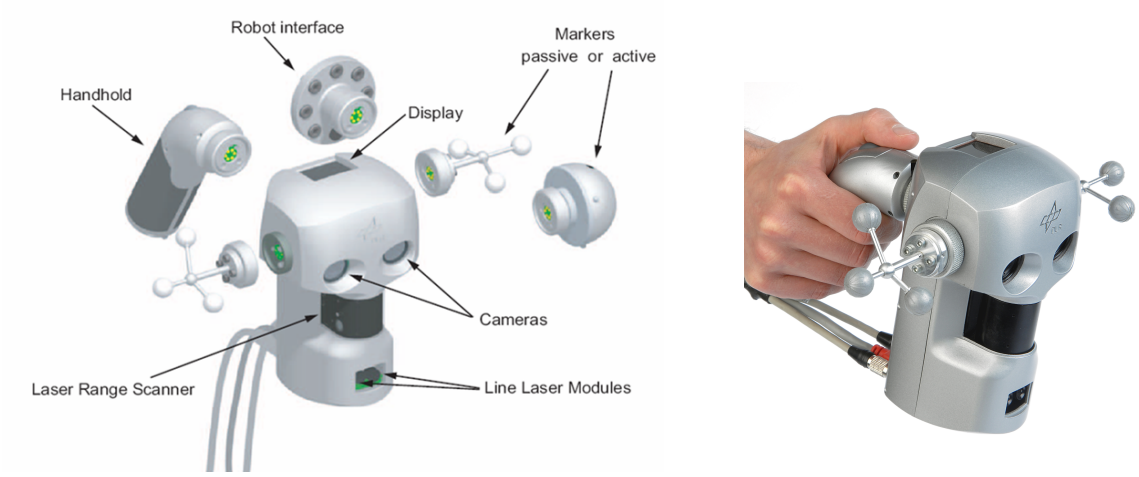

Figure 1: The DLR multi-sensory 3D-Modeler and its components.

measurement units as described in detail in [SKL+07], i.e., a laser-range scanner, a laser-stripe profiler, and a stereo depth-sensor. Furthermore, special mounts on the housing allow to use different external pose measurement units such as optical tracking systems, passive pose measurement arms, or a robot arm.

3.2 Synchronisation

Merging data from multiple sensors is achieved by a combined strategy of hardware and software synchronisation [BSSH07]. Here, synchronisation of a component is either obtained through periodic, electrical synchronisation pulses (h-synced) or through periodic software messages on a deterministic bus (s-synced) as shown in Figure 2.

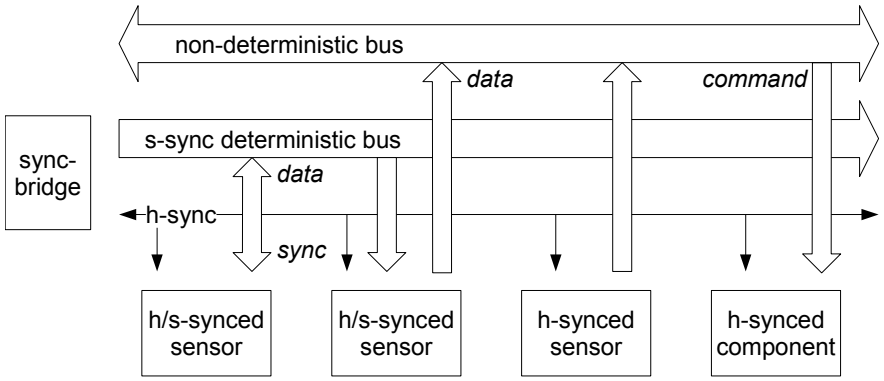

Figure 2: Synchronisation and communication concept.

All sensor data are labeled with a global timestamp contained in the software synchronisation messages before they are transmitted over any type of communication channel. In this manner, the system allows the data to be processed asynchronously and with arbitrary latencies. The label guarantees that finally the individual processing paths can be merged consistently.

3.3 Calibration

All three range-sensor units in the hand-held device need to be calibrated intrinsically and extrinsically in order to allow for a geometrically correct 3D reconstruction.

The stereo-camera unit is calibrated in a two-step procedure using the hand-eye calibration suite DLR CalDe/CalLab[4]. The intrinsic parameters of the stereo-camera pair are determined using a calibration pattern of known spatial dimensions. In a second step, the extrinsic rigid-body transformation is determined that refers the sensor coordinate system to the coordinate system measured by the external pose sensing device [SH06], e.g., the optical tracking system.

The laser-range-scanner is calibrated intrinsically with a coordinate measurement machine. The extrinsictransformation is determined by measuring a sphere of known extensions but unknown location with the hand-held device [SH04].

The extrinsic transformation of the laser-stripe profiler corresponds to the one of the stereo-camera system. Accordingly, only the 3-DoF location of the laser plane(s) with respect to the cameras is determined in a calibration procedure. This problem is solved referring taken measurements to a geometrically known object, in this case a planar surface [SSW+04].

[4] http://www.robotic.dlr.de/callab/

4 Streaming Generation of 3D Models

For the 3D-Modeler system, the in-the-loop generation of a 3D surface model for visual feedback is tackled by a streaming surface-reconstruction approach [BH04]. A dense and homogeneous triangular mesh is generated incrementally and from the real-time stream of 3D measurements. Since the generated model is available at any time instant, the data can be used in a real-time visualisation. The surface reconstruction algorithm consists of four processing stages, as shown in Figure 3.

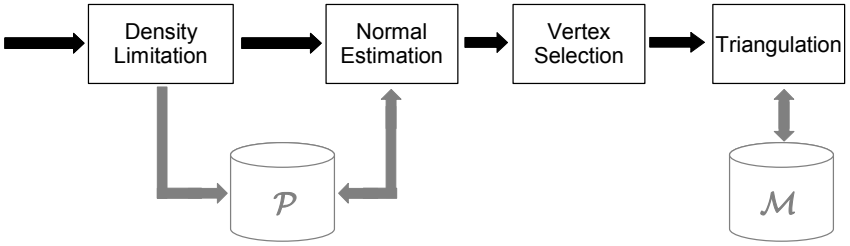

Figure 3: Pipeline for 3D surface-reconstruction from unordered, streamed 3D-point data. Local data storages are the accumulated point set $\mathcal{P}$ and the set of triangulated vertices $\mathcal{M}$.

4.1 Density Limitation

By manually scanning an object, multiple overlapping sweeps with the scanner generate spots of high local point density in the accumulated point set. The redundancies in the data increase the overall calculation effort without improving the result. For this reason, the density of the accumulated sample points is limited before any further processing. Since no shape information is available at this stage, the density can only be determined in the volume and not on a surface. Accordingly, the Euclidean distance between the measured points is used.

4.2 Normal Estimation

The surface normal for each sample point is the first local surface property that is derived from the unorganised point set and that is required in subsequent stages: Every time a new measurement triggers the normal estimation stage, all points inside a spherical volume around this sample are determined. For each of those points, the surface normal is determined by fitting a tangent plane through a neighbourhood of the examined point.

4.3 Vertex Selection

A selection denotes the verification of all modified surface normals after estimation. During mesh generation, a point with a miss-estimated surface normal potentially leads to cracks in the generated surface. Hence, points with bad surface normals are retained and only points with correct normals are passed to the triangulation stage.

4.4 Triangulation

The triangulation stage incrementally generates a homogeneous triangle mesh by continuously extending and refining the existing surface model with every newly inserted point. Therefore, each point that passes the selection stage is inserted as vertex into the existing triangle mesh and a localised 2D re-triangulation is performed in order to integrate the new vertex using the previously estimated surface normal.

5 Acquisition and Mapping of Textures

The acquisition and mapping of textures is an important step in the generation of photo-realistic 3D models. The 3D-Modeler is equipped with two colour-cameras with which the texture of the real object can be captured. Usually, multiple texture candidates (views) exist for a single surface patch. As described in [HBH+05], three texture-mapping algorithms are considered:

- single-view mapping,
- single-view mapping with brightness correction, and
- multi-view mapping

The first method chooses texture patch with the highest resolution among all captured patches. In dependence of the surface reflectance and environmental lighting, a face of the 3D model can appear in different brightness levels with varying viewpoints. In order to compensate for these effects, the latter two texture-mapping methods have been considered.

The second method still selects the texture patch with the highest resolution for a single face. By contrast, however, the brightnesses of all camera views are changed so that variations in commonly seen surface patches are minimised.

The third method merges texture information from multiple-views on the the surface patch. The method extracts colour information in all texture patches at the dominating resolution. These values are weighted according to the resolution of the originating texture patch.

As pointed out in [HBH+05], the texture images and 3D model of the object can be acquired either contemporaneously or sequentially to the scanning process. In order to prevent shadows on the surface, manual acquisition of texture images is preferred. Nevertheless, the operator is still requested to pay attention to shadows casting the object. In addition, the capture parameters such as shutter, gain and white balance are held constant to assure unchanging perception of surface brightness and colour.

6 A Mixed-Reality Visualisation for Hand-guided 3D Scanning

The 3D-Modeler, its synchronisation and communication infrastructure, as well as the streaming surface-generation services and texture mapping capabilities allow for a mixed-reality system for the intuitive hand-guided digitisation of small-scale and mid-scale objects.

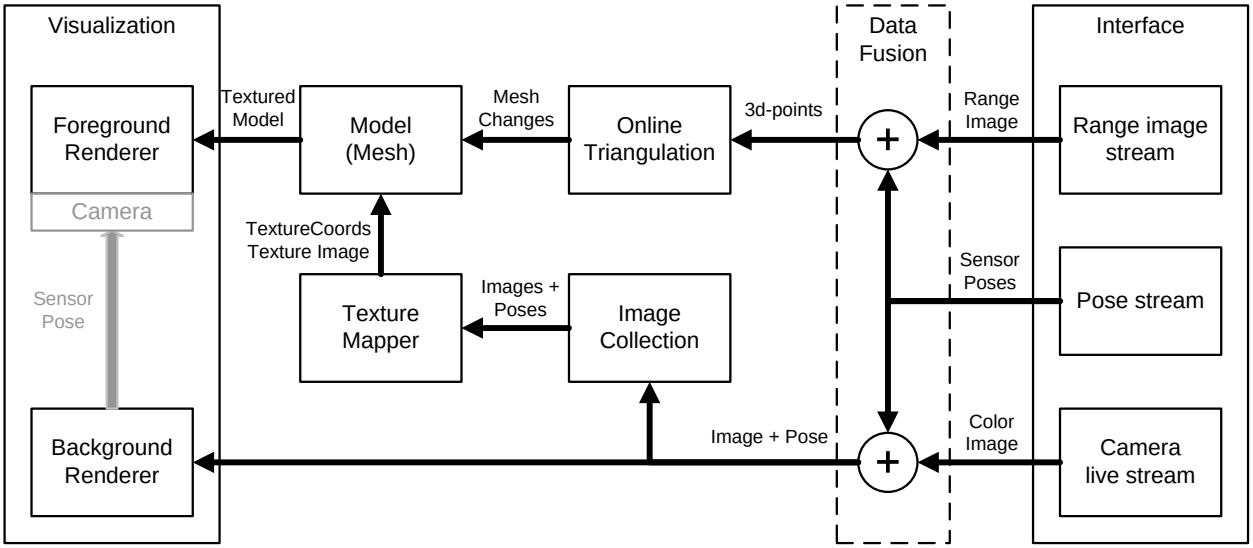

Figure 4: Data flow for mixed-reality visualisation.

6.1 Synchronisation and Latency

The data-streams of range sensors and the colour camera are consistently merged with the data from the pose-sensor following the synchronisation approach of Section 3.2 (see Figure 4).

In our system, the frequency of the global clock is 25 Hz. Accordingly, synchroneity is assured within one period, i.e., 40ms. Nevertheless, latencies differ from sensor to sensor and vary due to transmission over non-deterministic busses and due to data processing on non-real-time operation systems. Table 1 shows the expected latencies for individual data streams. Buffers of reasonable sizes are used to deal with the problem in the data-merging module.

The empirically set buffer sizes cope with different system configurations. Different external pose measurement devices are considered, e.g., the optical tracking system ART2 (http://www.ar-tracking.de) based on retro-reflecting markers, the robot LWR3 [HSF+09], and an image-based ego-motion estimation [SMB+09].

With respect to range measurements, the 3D-Modeler offers three sensors. All of these sensors show different latencies. In general, all camera-based information shows the highest latencies.

measurement type		latency		
		sensing	transmission	processing
depth	rotating laser scanner	30ms	-	-
	laser-stripe profiler	5ms	40ms	20ms
	correlation-based stereo (320x240)	5ms	40ms	400ms
pose	ART smARTtrack2	25ms	-	-
	ego-motion estimation	5ms	40ms	40ms
texture	colour camera (780x580)	5ms	40ms	40ms

Table 1: Maximal latencies of individual data-streams on a CoreDuo 3.0 GHz.

Table 1 refers to the data collection on the machine that hosts all interfaces to the sensor devices. If the data is merged on remote machines, then an additional latency for RPC-TCP/IP communication has to be taken into account.

6.2 View Planning

The view planning is supported by an immediate visualisation of the scanning progress. We distinguish four different types of visualisations by a so-called seamless factor as reported in Table 2. The closer the rendered information is to how the operator actually sees the real object, the higher is the seamless factor.

seamless factor	visualisation modality
high	textured surface model
	shaded surface model
	wire-frame model
low	raw point model

Table 2: Modalities of surface visualisation that support view planning.

seamless factor	view alignment by	display
high	operator pose	see-through HMD
	sensor pose	desktop
	measurement	desktop
low	mouse	desktop

Table 3: Possible modes of view alignment supporting the navigation task.

The simplest visualisation provides the operator only with the information of the sensing unit. More valuable is the immediate filtering and triangulation of 3D point measurements as described in Section 4 combined with the immediate visualisation of these results. Accordingly, the operator is able see the problems associated to the measurement or to triangulation while scanning. Moreover, the visualisation of the triangulation result supports the operator in the navigation task because the virtual information better reflects the real world.

At the final stage, and as soon as a sufficient geometric reconstruction of the object exists, radiometric information is gathered as described in Section 5. Figure 5 shows the different visualisation modes while digitising a bust of an native American.

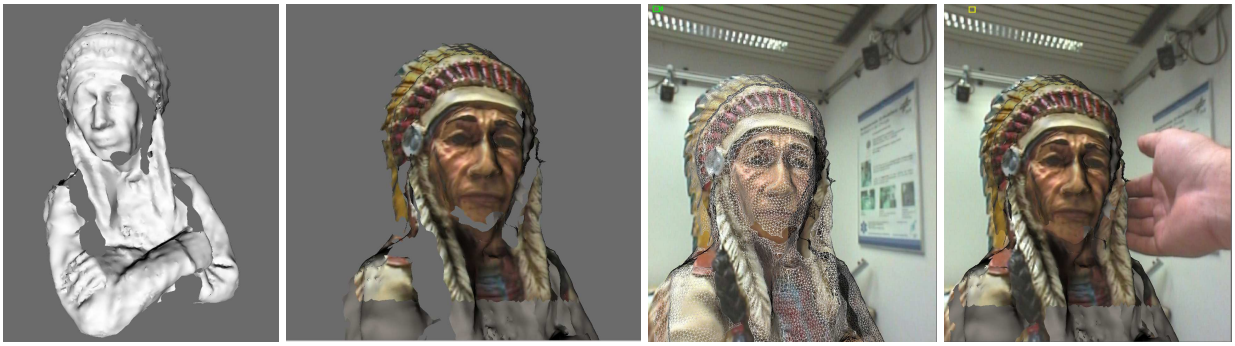

Figure 5: Different modalities of visualisation. From left to right: surface-model, textured surface-model, mixed-reality wire-frame model, mixed-reality textured model.

6.3 Navigation

The operator requires an appropriate view to the 3D model that reflects his focus and that allows him to easily localise the shown parts on the real object. However, the ”appropriate view” is subjective, i.e., it is different for every user. We implemented three ways to set the viewpoint as shown in Table 3. Again, we distinguish according to a seamless factor that reflects the closeness of the operator's viewpoint to the rendered scene.

The simplest mode allows the operator focus on details in the scene by setting the view of the virtual camera via mouse and keyboard. However, he has to switch between the scanner device and the mouse to change the view.

When scanning surfaces continously, it is more appropriate to dynamically change the view according to the currently processed surface parts or the scanner movement. A view alignment w.r.t. the last processed surface part is implemented by setting the camera view to the viewpoint of the respective depth image.

With this approach, however, the visualisation view is only updated while scanning the surface. The availability of continuous pose measurements of the hand-held device allows the use of the scanner itself as a 6-DoF input device. Accordingly, the user can inspect the 3D model in the virtual world by moving the scanner around the real object.

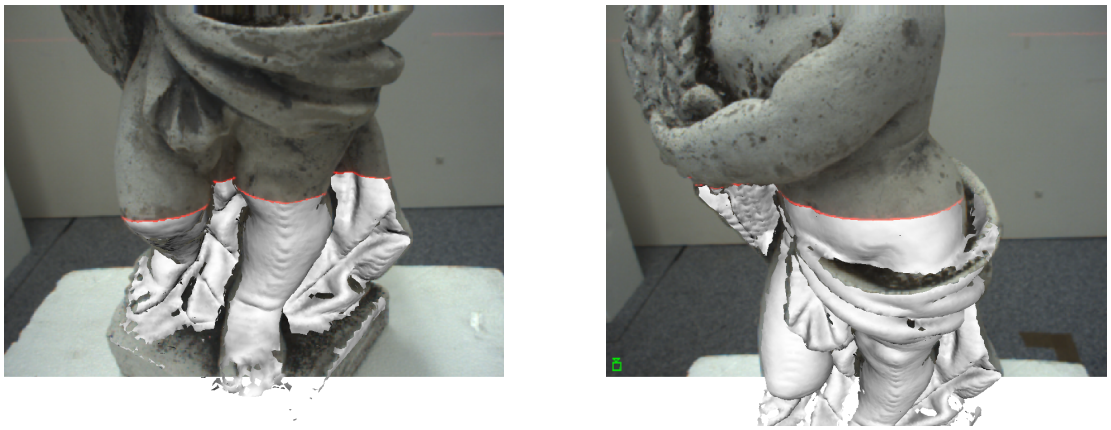

Figure 6: Mixed-reality immediate visual-feedback during 3D-scanning operation.

Recognising the surface parts on the display is still tedious, especially at the beginning of the scanning process. The visualisation can further support the user at this task by augmenting the 3D model with a live video-stream from a camera of the hand-held device as shown in Figure 6. The generated 3D model is shown in front of the live stream at the pose that matches the camera view. This image-embedded model helps to match the virtual 3D model with the real object and helps to navigate over the surface with the device. Moreover, the live video-stream mode allows to take appropriate snapshots for texture mapping (see Figure 7).

The rendered 3D model seamlessly fits the real object shown in the live camera-image only if the camera parameters of video stream correspond to the parameters of the virtual camera. Here, we undistort the images from the video stream and use the remaining intrinsic parameters for

setting the perspective projection of the virtual camera.

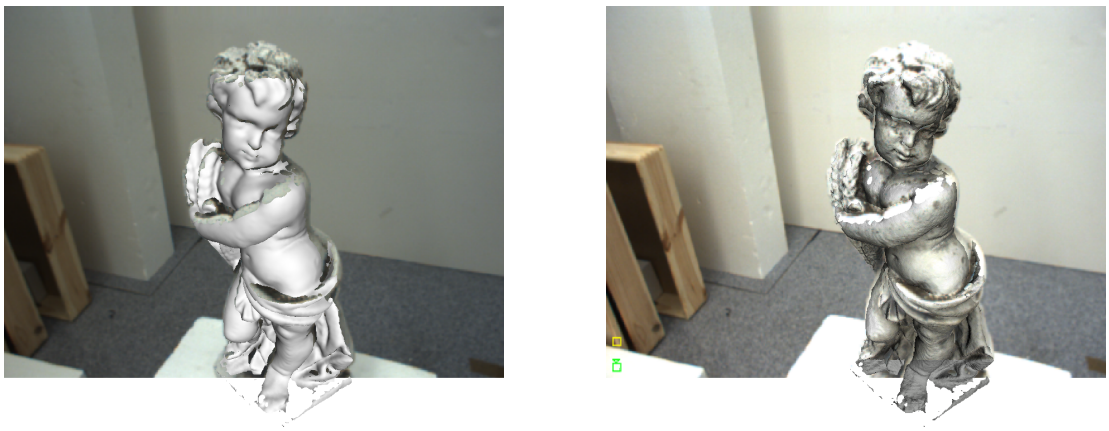

Figure 7: Mixed-reality visualisation. Left: untextured model. Right: textured model.

7 Conclusion

We present an approach that supplies the operator of the hand-held scanning device with immediate visual-feedback. This feedback is significantly enhanced using a mixed-reality visualisation of the virtual 3D-model and the real object gathering synchronised images from embedded cameras.

Such an approach efficiently supports the operator in two main aspects of the scanning procedure, i.e., view planning and navigation. The paper addresses the requirements for immediate data processing and visualisation. Further, it summarises the hardware and software components of a suitable system, the 3D-Modeler.

A future combination of a mixed-reality visualisation and a head-mounted see-through display could support the operator in his task even better. However, it is yet to show that the spatial and temporal alignments in this visualisation are accurate enough to meet the operator's need.

Acknowledgements

The authors would like to thank Ulrich Hagn, Simon Kielhöfer, and Franz Hacker for hardware support as well as Klaus Strobl, Rui Liu, and Stefan Fuchs for software support.

References

[AST+03] P. K. Allen, I. Stamos, A. Troccoli, B. Smith, M. Leordeanu, and Y. C. Hsu. 3d modeling of historic sites using range and image data. In *Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA*, pages 145–150, Taipei, Taiwan, 2003.

[BH04] T. Bodenmüller and G. Hirzinger. Online surface reconstruction from unorganized 3d-points for the dlr hand-guided scanner system. In *2nd Symp. on 3D Data Processing, Visualization, Transmission*, pages 285–292, Thessaloniki, Greece, 2004.

[BSSH07] T. Bodenmüller, W. Sepp, M. Suppa, and G. Hirzinger. Tackling multi-sensory 3d data acquisition and fusion. In *Proc. of the IEEE/RSJ Int. Conf. on Intell. Robots and Systems, IROS*, pages 2180–2185, San Diego, CA, USA, 2007.

[FFG+96] R. Fisher, A. Fitzgibbon, A. Gionis, M. Wright, and D. Eggert. A hand-held optical surface scanner for environmental modeling and virtual reality. In *Proc. of Virtual Reality World*, pages 13–15, Stuttgart, Germany, 1996.

[HBH+05] G. Hirzinger, T. Bodenmüller, H. Hirschmüller, R. Liu, W. Sepp, M. Suppa, T. Abmayr, and B. Strackenbrock. Photo-realistic 3d modelling – from robotics perception towards cultural heritage. In *Int. Workshop on Recording, Modeling and Visualization of Cultural Heritage 2005*, Ascona, Italy, 2005.

[HI00] A. Hilton and J. Illingworth. Geometric fusion for a hand-held 3d sensor. In *Machine Vision and Applications*, volume 12, pages 44–51, Springer-Verlag, 2000.

[HSF+09] S. Haddadin, M. Suppa, S. Fuchs, T. Bodenmüller, A. Albu-Schffer, and G. Hirzinger. Towards the robotic co-worker. In *14th Int. Symp. on Robotics Research, ISRR*, Lucerne, Switzerland, 2009.

[SA00] I. Stamos and P. K. Allen. 3-d model construction using range and image data. In *Proc. of the IEEE Conf. on Comp. Vision and Pattern Recog., CVPR*, pages 531–536, 2000.

[SH04] M. Suppa and G. Hirzinger. A novel system approach to multisensory data acquisition. In *The 8th Conf. on Intell. Autonomous Systems, IAS-8*, pages 996–1004, Amsterdam, The Netherlands, 2004.

[SH06] K. H. Strobl and G. Hirzinger. Optimal hand-eye calibration. In *Proc. of the IEEE/RSJ Int. Conf. on Intell. Robots and Systems, IROS*, pages 4647–4653, Beijing, China, Oct 9–15 2006.

[SKL+07] M. Suppa, S. Kielhöfer, J. Langwald, F. Hacker, K. H. Strobl, and G. Hirzinger. The 3d-modeller: A multi-purpose vision platform. In *Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA*, pages 781–787, Rome, Italy, Apr 10-14 2007. IEEE.

[SMB+09] K. H. Strobl, E. Mair, T. Bodenmüller, S. Kielhöfer, W. Sepp, M. Suppa, D. Burschka, and G. Hirzinger. The self-referenced dlr 3d-modeler. In *Proc. of the IEEE/RSJ Int. Conf. on Intell. Robots and Systems, IROS*, page in press, St. Louis, MO, USA, 2009.

[SSW+04] K. H. Strobl, W. Sepp, E. Wahl, T. Bodenmüller, M. Suppa, J. F. Seara, and G. Hirzinger. The DLR multisensory hand-guided device: The laser stripe profiler. In *Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA*, pages 1927–1932, New Orleans, LA, USA, 2004.