Real-Time Active Multiview 3D Reconstruction

Kai Ide
Communication Systems Group
Technische Universität Berlin
10587 Berlin, Germany
Email: ide@nue.tu-berlin.de

Thomas Sikora
Communication Systems Group
Technische Universität Berlin
10587 Berlin, Germany
Email: sikora@nue.tu-berlin.de

***Abstract*—We report on an active multiview system for real-time 3D reconstruction based on phase measuring triangulation. Our system overcomes one of the greatest drawbacks in active 3D reconstruction, namely occlusions due do shadowing of either the camera or the projector light source. Our system is comprised of two high speed cameras in conjunction with two projectors and is currently capable of capturing and rendering up to 5.2 million 3D vertices at 10 fps. Four additional color cameras provide for texturing the underlying 3D geometry, thus making the system suitable for real-time view synthesis on conventional, stereoscopic or novel autostereoscopic multiview displays.**

***Keywords*—3D reconstruction, 3D scanning, structured light, phase shifting, 2D phase unwrapping, free viewpoint video.**

I. Introduction

Everything around us we see in color and in 3D. Stereoscopic television provides a more realistic experience than monoscopic television but real-world motion parallax, that would allow viewers to see behind objects by moving their head, is not provided for. Additionally, capturing stereoscopic content has proven to be challenging since it requires recording not one but two images with cameras that capture in perfect synchrony, exhibit identical colorimetric properties, have identical focal lengths, apertures, field of depth and so on. Today, this can be achieved by skilled stereographers with modern equipment but the demand for more realism and for being able to view 3D content without special eyewear has driven the development of autostereoscopic multiview displays. Multiview displays, to a certain extend, provide for horizontal motion parallax. These displays require five to thirty input views, all with the same high degree of image alignment as mentioned above. Volumetric or holographic displays provide for full motion parallax. Viewers can freely move around in order to get new perspectives onto the scene. However, volumetric and especially holographic displays require a number of input views that is even orders of magnitude higher. Recording such imagery quickly becomes infeasible. Computer Generated Imagery (CGI) mitigates an ample amount of the challenges that exist when such image pairs are to be created. Given a geometric representation of a scene, CGI can render to any number of *virtual* cameras with perfectly matched properties. Real-time rendering capabilities provided, this allows to create interactive, full parallax Free Viewpoint Video (FVV) [1], [2]. It is this possibility to capture *and render* a 3D scene that opens up a host of possibilities across a variety of applications, ranging from CAD modeling of real world objects, to surface inspection and volumetric or even holographic rendering.

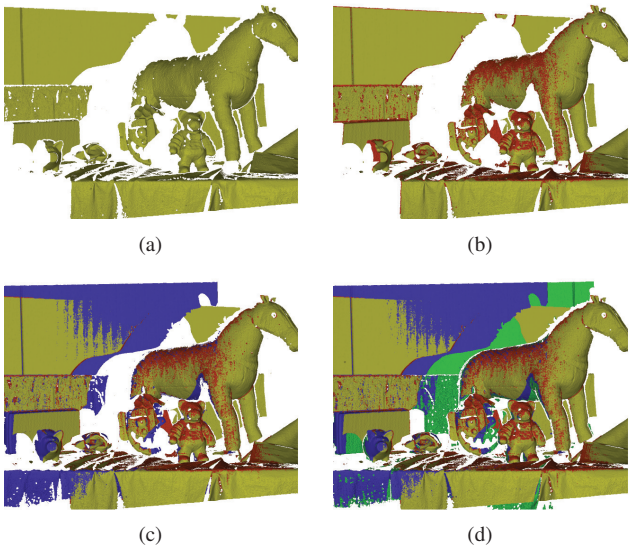


Figure 1. Color-coded multiview reconstruction illustrating the resulting gain in terms of reconstruction completeness with one (a), two (b), three (c), and four (c) 3D scanning units.

Ideally, a 3D camera should thus capture the 3D geometry of a scene along with its texture and reflective properties. For this reason we have designed a system able to capture time-varying 3D geometry in real-time, that is both as complete and as accurate as possible. Our setup is designed to capture time varying 3D geometry within a relatively large working volume of approximately 2.5 m×2.5 m×3.0 m.

II. Related Work

Image-based reconstruction can roughly be divided in two categories, namely *active* and *passive* techniques [3], [4]. In the past years, both techniques have emerged as commercial systems that perform real-time geometry acquisition. Available systems include but are not limited to the passive trifocal Point Grey Bumblebee XB3 camera, PMD[vision]'s active CamCube 3.0 Time of Flight (ToF) camera, and consumer grade systems such as Microsoft's Kinect sensor, which also falls within the category of active techniques.

Passive 3D reconstruction techniques rely solely on ambient light and, apart from Structure from Motion (SfM) or Depth from Defocus (DfD) approaches, require two or more cameras. Passive techniques suitable for real-time reconstruction at

video frame rates include [5], where local stereo is performed using the concept of guided image filtering demonstrated in [6]. Hybrid recursive stereo matching using a trifocal plus a wide baseline stereo rig has been demonstrated in [7]. The method has recently been extended to allow for spatio-temporal consistency [8]. A concise overview of the state of the art in passive stereo is given in [9] and most notably in the Middlebury dataset [10]. Development towards multiview reconstruction is accounted for as a related benchmark in [11].

Active reconstruction is in contrast generally more accurate. Techniques utilizing structured light (SL) perform 3D reconstruction by measuring the deformation of a known light pattern after it is reflected from the scene and captured with a camera [12]. A concise summary of structured light patterns is given by Salvi et al. [13]. In the field of 3D reconstruction, Phase Shifting (PS) or Phase Measuring Triangulation (PMT) is considered the state of the art in terms of accuracy and data density. Requiring a sequence of at least three phase shifted sinusoidal patterns, reconstruction via PMT has been demonstrated in [14]. Extending the method to allow for motion compensation and the scanning of discontinuous objects, Weise et al. [15] perform 3D scanning in real-time, by utilizing a projector and a stereo camera rig for phase unwrapping. A four pattern phase shift with an integrated binary coding scheme based on DeBruijn sequences is applied by Wissmann et al. in [16]. Utilizing the binary coding scheme for phase unwrapping, real-time 3D reconstruction at 6 – 11 fps is demonstrated. A common drawback in the design of these systems is that they are build to capture 3D geometry within a relatively small working volume, thus confining their application to head scanning or 3D reconstruction of relatively small and rigid objects. Additionally, these systems suffer from shadowing since they utilize only a single light source.

III. System Overview

A. Hardware setup

Our system is comprised of two Viewsonic 120 Hz PJD6241 DLP projectors and two Basler A504k SXGA cameras, that capture at a maximum framerate of 500 fps. Four Basler Scout scA1300-32gc 1.3 megapixel color cameras provide texture information. Two of these, along with one projector and one high speed camera form a scan unit, as depicted in Fig. 2. Configuring the scan units to capture SL from both projectors returns two additional wide base line camera-projector pairs. This yields a total of four pairs which capture simultaneously, each contributing to the 3D reconstruction as shown in Fig. 1. The video input to the first projector P_1 is analyzed by an external Atmel EVK1100 AVR32 microcontroller in order to maintain synchronization with the SL sequence and providing the triggering timings to the camera array. The SL sequence itself is projected by a dedicated *secondary* computer C_2, while a *primary* computer C_1 does all of the 3D processing. A TCP/IP socket and a USART interface serve as communication links to C_2 and the microcontroller, respectively.

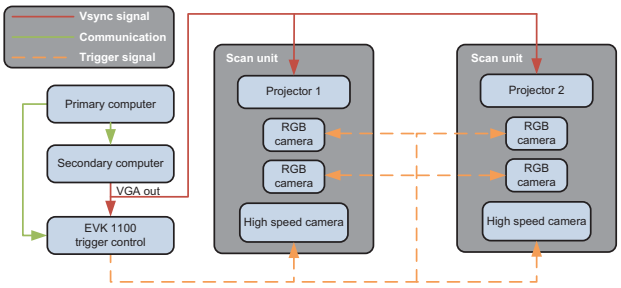


Figure 2. Active 3D reconstruction system layout.

B. Camera projector synchronization

Synchronization among the projection sequence and the triggering sequence is maintained by continuously analyzing the video signal fed to the first projector P_1. By distinguishing whether the current frame is active or not over several frames, we can exactly derive our position within the projection sequence. To reduce the impact of noise we limit our integration time to the interval $[t_0, t_1]$, with a delay of $t_0 = 2.1\,\text{ms}$ w.r.t the occurrence of a VSync signal. Integration ends at $t_1 = {}^1\!/_{120\,\text{Hz}} = 8.33\,\text{ms}$. A binary decision is then made by evaluating the green channel v_g of VGA input to P_1 with $v_g = \int_{t_0}^{t_1} g(t)\mathrm{d}t$ and thresholding this against $\frac{\max(v_g)-\min(v_g)}{2}$.

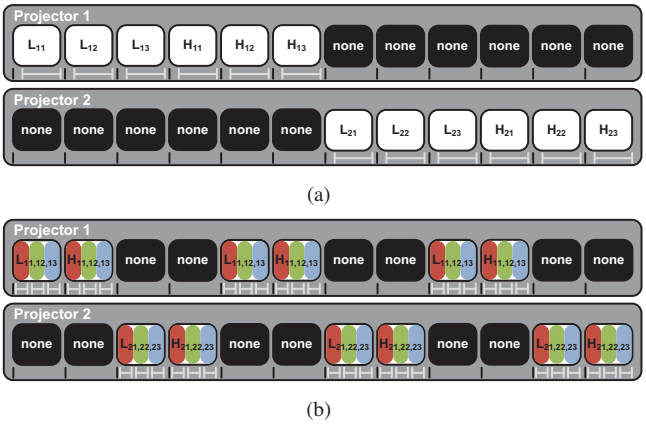
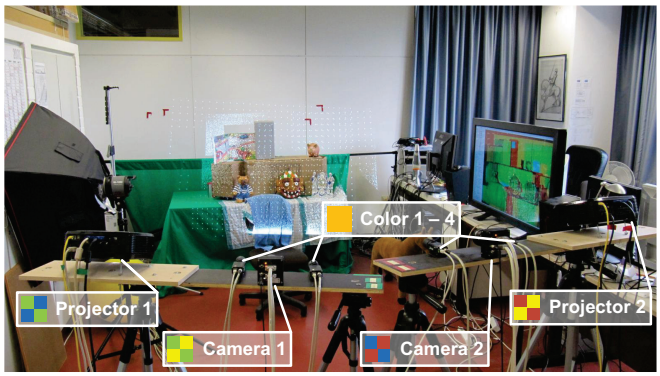


Figure 3. The time multiplex schemes of our structured light sequence, running at 10 fps (a) and 30 fps (b). VSync signals are indicated by black ticks whereas the high speed camera's acquisition intervals are denoted by gray bars.

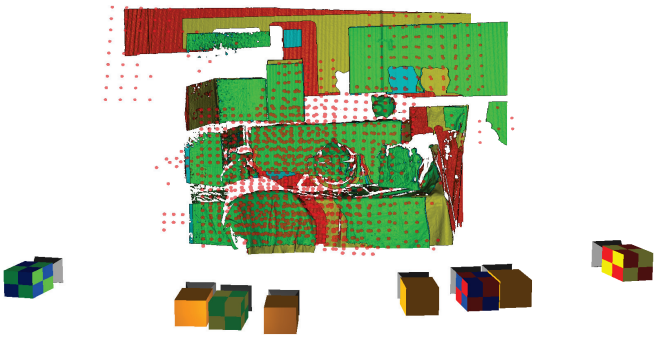
IV. System Calibration

A. Geometric multi-camera multi-projector calibration

The geometric calibration of the multi-camera multi-projector rig is done fully automatically with a technique that we have described in [17]. By sequentially projecting pointclouds of binary coded markers into the scene with each projector and capturing the individual bits of this point cloud with all cameras on a frame by frame basis, we can derive a dense and error free pixel correspondence set among all the cameras and projectors. This set forms a highly overdetermined system of linear equations which is solved w.r.t



(a)



(b)

Figure 4. Physical system layout (a) and the result of automatic multi-camera multi-projector calibration (b). The calibration pointcloud (red) is shown on top of the colorcoded 3D reconstruction of the four camera-projector pairs.

to the unknown camera parameters [18]. After the feature point correspondence matrix has been processed, the setup is calibrated within a world coordinate system W^* centered in the middle of the 3D pointcloud. The three major principal components of the 3D pointcloud form the basis vectors of this coordinate system. In order to arrive at a new world coordinate system W which does not change its center, orientation and scale w.r.t to the physical camera setup upon every new calibration, we normalize all extrinsic parameters, so that the world coordinate system coincides with a reference camera.

Let r denote a reference camera, and let $\mathbf{N} = \mathbf{C}^*_\mathbf{r} - \mathbf{C}^*_{\mathbf{r+1}}$ be the vector between the center of r and the center of camera $r+1$, with the camera centers given by

$$\mathbf{C}^*_\mathbf{i} = -\mathbf{R}^{*\prime}_\mathbf{i}\mathbf{t}^*_\mathbf{i} \tag{1}$$

This yields a normalization parameter n, that is applied for scaling the new extrinsic camera parameters

$$n = \frac{\mathbf{N}}{s \cdot |\mathbf{N}|} \tag{2}$$

The extrinsic translation vectors are then given by

$$\mathbf{t_i} = -\mathbf{R}^*_\mathbf{i}\frac{(\mathbf{C}^*_\mathbf{i} - \mathbf{C}^*_\mathbf{r})}{n}, \tag{3}$$

where the corresponding extrinsic rotation matrices are written as

$$\mathbf{R_i} = \mathbf{R}^*_\mathbf{i}\mathbf{R}^{*\prime}_\mathbf{r} \tag{4}$$

A transformation matrix $\mathbf{M_i}$ is then constructed from $\mathbf{R_r}$ and $\mathbf{t_r}$, warping each homogeneous vertex $\mathbf{X}$ in the 3D pointcloud into the new world coordinate system

$$\mathbf{X} = \mathbf{M_i}\mathbf{X}^* = \begin{bmatrix} \mathbf{R_r} & \mathbf{t_r}/n \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{X}^* \tag{5}$$

This allows for a comparison between the physical setup and the calibration data, as illustrated in Fig. 4.

B. Gamma inversion

As a prerequisite step, projection of a sinusoidal pattern requires a linearization of the projector's response to a linear ramp function $R(x) = x$ with $x \in [0, \ldots, 255]$, x representing overall image intensity. Partially, this nonlinearity results from gamma correction in the secondary computer C_2 *and* the projectors. The cameras are assumed to be linear. C_2's operating system's gamma value is $\gamma = 2.2$ but directly calculating the inverse with $I_{\text{out}} = I_{\text{in}}^{-\gamma}$ results in an imperfect reconstruction. Thus, the response is measured several times yielding its average inverse in the form of a lookup table. The measured gamma curves and their inverses vary slightly among the projectors as illustrated in Fig. 5.

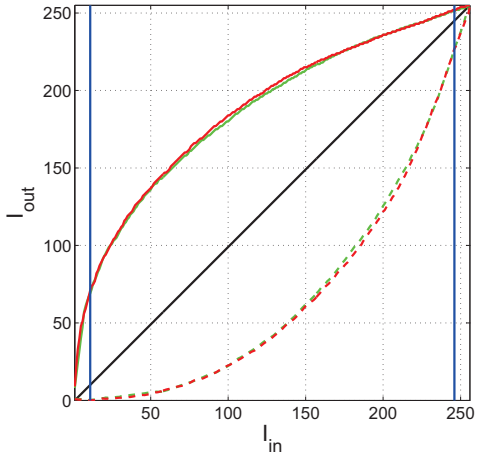


Figure 5. Each projector's response to the intensity ramp function $R(x)$, yielding the averaged dashed gamma curves and their continuous inverses (P_1 : red, P_2: green). The usable intensity range has been reduced to the interval $[10, 245]$, as indicated by the blue vertical lines.

V. 3D RECONSTRUCTION

Our system is optimized for fast 3D reconstruction, thus it is noteworthy to mention that all of the later image processing steps required for 3D triangulation are performed on a per pixel basis. Our method does not require searching for pixel to pixel correspondence but incorporates all the necessary 3D

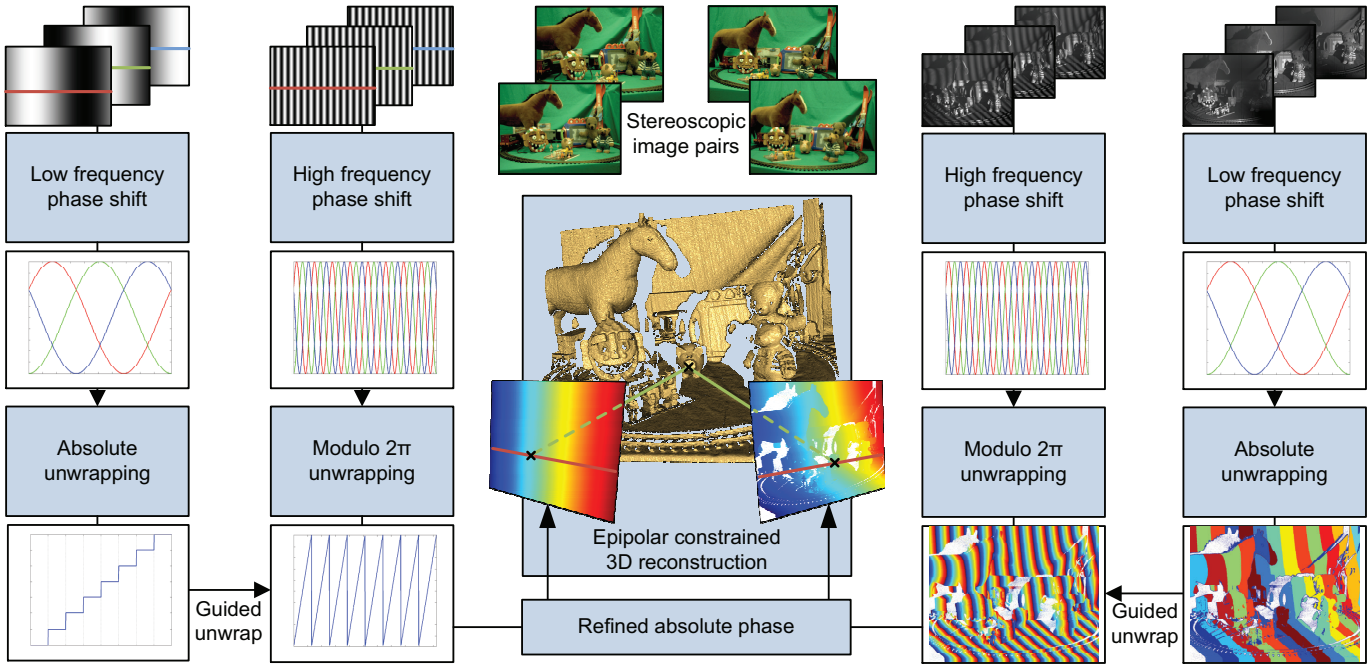


Figure 6. Our phase unwrapping procedure for active 3D scene reconstruction, showing corresponding projected images (left) and captured images (right). A high frequency phase shift triplet is converted to a modulo 2π wrapped phase. Absolute phase recovery is performed by utilizing a low frequency phase shift triplet for guided phase unwrapping. Epipolar constrained 3D reconstruction is shown in the center, stereoscopic color image pairs (above) provide additional texture.

information within the structured light sequence. Because of this, we are able to reconstruct up to 52 million 3D vertices per second, which in turn makes it feasible to utilize information from multiple 3D scan units simultaneously. All of the 3D processing runs on the GPU in parallel. The workload on behalf of the CPU is limited to merely capturing and uploading image data to the GPU.

We sequentially project two phase shifted sinusoidal triplets - one of low frequency L_{ij} and one of high frequency H_{ij}. The later is converted to a modulo 2π wrapped phase. Absolute phase recovery is performed by utilizing the low frequency patterns for guided phase unwrapping. Epipolar constrained triangulation yields the final 3D reconstruction, while the stereoscopic color image pairs provide additional texture. Our technique is summarized in Fig. 6.

As stated before, two vertically invariant phase shifted sinusoidal triplets are emitted with each projector. The patterns are given by:

$$\begin{aligned} p_1(x) &= I_{dc} + I_a * \cos(\Phi - \Theta) \\ p_2(x) &= I_{dc} + I_a * \cos(\Phi) \\ p_3(x) &= I_{dc} + I_a * \cos(\Phi + \Theta) \end{aligned} \tag{6}$$

The image's mean is denoted by I_{dc}, I_a is the signal's modulation amplitude, Φ is the phase across the images given by $\Phi = {}^{2\pi N x}/_{w}$, where the projector's horizontal resolution corresponds to w with the pixel indices on it given by x. Θ is the actual phase shift of $\Theta = {}^{2\pi}/_{3}$. In the case of the high frequency triplet, a number of periods with $N = 32$ was selected. In the case of the low frequency triplet we only project a single period with $N = 1$.

A. Phase unwrapping

Having captured a complete phase shifted triplet we can derive its underlying phase, normalized between $[0, 1]$, with:

$$\Phi'(x) = \frac{\arctan\left(\frac{\sqrt{3}(p_1 - p_3)}{2p_2 - p_1 - p_3}\right) + \pi}{2\pi} \tag{7}$$

The low frequency phase then guides the phase unwrapping of the high frequency triplet, the phase of which is modulo 2π wrapped. This is performed on a per pixel basis by evaluating all N possible locations within the phase. The absolute phase is then unwrapped by:

$$\Phi(\mathbf{x}) = \frac{1}{N}\left(\Phi'_h + \lfloor N\Phi'_l + 0.5 \rfloor\right) \tag{8}$$

In practice it is beneficial to apply a denoising filter to Φ'_l to remove outliers. This is achieved by bilateral filtering within a small $[3 \times 3]$ neighborhood $\mathbf{p}$ around pixel $\mathbf{x}$

$$\Phi^*_l(\mathbf{x}) = \frac{1}{N}\sum_{\forall p_x}\sum_{\forall p_y} G_{\sigma_d}(\mathbf{p} - \mathbf{x})\, G_{\sigma_r}(\Phi'_l(\mathbf{p}) - \Phi'_l(\mathbf{x}))\, \Phi'_l(\mathbf{x}) \tag{9}$$

with the Gaussian filters

$$G_{\sigma_d}(\mathbf{p}-\mathbf{x}) = \frac{\exp\left(-\frac{(p_x-x_x)^2+(p_y-x_y)^2}{2\sigma_d^2}\right)}{2\pi\sigma_d^2}, \quad (10)$$

$$G_{\sigma_r}\left(\Phi_l'(\mathbf{p})-\Phi_l'(\mathbf{x})\right) = \exp\left(-\frac{\left(\Phi_l'(p_x,p_y)-\Phi_l'(x_x,x_y)\right)^2}{\sigma_r^2}\right) \quad (11)$$

and a normalization term

$$N = \sum_{\forall p_x}\sum_{\forall p_y} G_{\sigma_d}\left(\mathbf{p}-\mathbf{x}\right) G_{\sigma_r}\left(\Phi_l'(\mathbf{p})-\Phi_l'(\mathbf{x})\right), \quad (12)$$

while the parameters for domain and range variance are set to $\sigma_d = 18$ and $\sigma_r = 0.07$, respectively.

B. Epipolar constrained triangulation

The absolute phase establishes a direct relationship between the homogeneous camera pixels $\mathbf{x_c}$ and the homogeneous projector pixels $\mathbf{x_p}$, which can be derived by calculating the corresponding epipolar line $\mathbf{l}$ in the projector's image plane

$$\mathbf{l} = \mathbf{F}\mathbf{x_c}, \quad (13)$$

where $\mathbf{F}$ denotes the fundamental matrix relationship between the camera and the projector. The pixel in the projector image plane is then given by:

$$\mathbf{x_p} = \begin{bmatrix} \theta \cdot w \\ \frac{\mathbf{x_{p_1}}\cdot\mathbf{l_1}+\mathbf{l_3}}{\mathbf{l_2}} \\ 1 \end{bmatrix} \quad (14)$$

Once a pixel to pixel correspondence has been established, a 3D vertex position can be triangulated. The limited computational complexity of the GLSL GPU environment calls for an estimation of a geometric ray-ray intersection, whereas conventionally a corresponding singular value decomposition, given two projection matrices and the pixel correspondence would suffice. The optical centers for camera $\mathbf{p}$ and projector $\mathbf{q}$ are calculated with Eq. 1. The rays formed by the camera / projector centers and the respective pixel locations within the image planes are then given by:

$$\mathbf{a}^* = \mathbf{R_c}'{\mathbf{K_c}}^{-1}\mathbf{x_c} \quad (15)$$

$$\mathbf{b}^* = \mathbf{R_p}'{\mathbf{K_p}}^{-1}\mathbf{x_p} \quad (16)$$

Normalization to unit length with $\mathbf{a} = {}^{\mathbf{a}^*}/_{|\mathbf{a}^*|}$ and $\mathbf{b} = {}^{\mathbf{b}^*}/_{|\mathbf{b}^*|}$ allows for the calculation of a vector perpendicular to the plane spanned by $\mathbf{a}$ and $\mathbf{b}$ and vector $\mathbf{b}$ itself:

$$\mathbf{t_a} = \mathbf{b} \times (\mathbf{a} \times \mathbf{b}) \quad (17)$$

The point of intersection of $\mathbf{a}$ with this plane is given by:

$$\mathbf{i_a} = \mathbf{p} + \frac{(\mathbf{q}-\mathbf{p})' \cdot \mathbf{t_a}}{\mathbf{a}\cdot\mathbf{t_a}} \cdot \mathbf{a} \quad (18)$$

Similarly, a corresponding point of intersection is found for $\mathbf{b}$, yielding $\mathbf{i_b}$. The resulting 3D vertex $\mathbf{x_w}$ is then given by:

$$\mathbf{x_w} = \frac{\mathbf{i_b}-\mathbf{i_a}}{2} + \mathbf{i_a} \quad (19)$$

Synthesis of novel views demands shadow mapping w.r.t. color cameras and projective texturing with $\mathbf{x_i} = \mathbf{P_i}\mathbf{x_w}$, where $\mathbf{x_i}$ denotes pixel coordinates and $\mathbf{P_i}$ represents the projection matrix of the i-th texturing camera. 3D meshes with per triangle normal vectors are precomputed as triangle strips, allowing for a shaded representation as shown in Fig. 7.

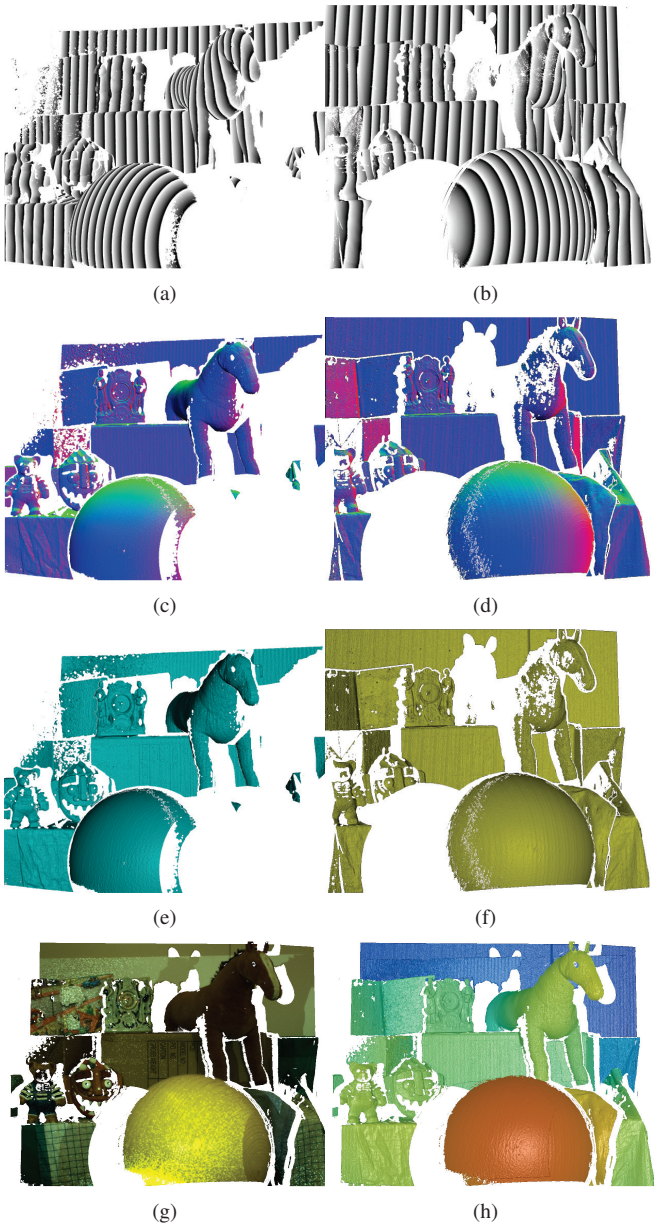


Figure 7. Modulo 2π phase (a) and (b), surface normal map (c) and (d), color coded partial reconstruction (e) and (f), novel textured view for combined reconstruction (g) and the corresponding perspective depthmap (h) (near: red, far: blue).

C. Timing overview

Timing diagrams for a complete reconstruction with the presented system, averaged over a sequence of 500 frames, are illustrated in Fig. 8. A complete multiview 3D reconstruction cycle accumulates to 76.5 ms including rendering, which currently manifests as the main bottleneck due to the high amount of generated 3D data. Image capturing from the cameras is performed by a dedicated CPU thread in parallel. Due to the demand to project and capture the entire SL sequence of 12 frames at 120 Hz, our system is currently limited to a maximum reconstruction frequency of 10 fps. The underlying main platform of C_1 is a quad core (3.2 GHz) i7 CPU, with 6 Gbyte of RAM and a Geforce GTX 295 GPU.

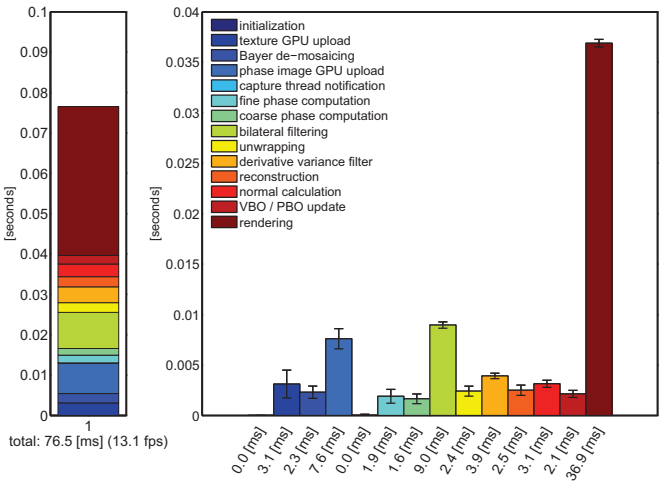


Figure 8. Averaged timing diagrams for a complete 3D scene reconstruction, showing the accumulative sum over all processing steps and their individual timing details.

VI. Conclusion

We have demonstrated active multiview 3D reconstruction based on phase measuring triangulation in real-time, running at 10 fps. Due to the utilization of multiple projectors and cameras we are able to greatly reduce the impact of shadowing and thus arrive at geometric 3D models with significantly less occlusions. The geometric 3D scene representation allows for the synthesis of stereoscopic and multiview content in real-time. Additional head-tracking equipment enables the interactive display of Free-Viewpoint Video. Future work will include projector lamp replacement with high power infrared emitters and the reduction of the sequence acquisition time, which is currently 47.8 ms, by compressing the phase shift into the RGB-channels of a single projection frame, as depicted in Fig. 3b. The relatively long sequence acquisition time additionally calls for motion compensation. In homogeneous regions the method described in [15] will be applied, in textured regions the compensation via dense optical flow fields appears promising. Additionally, we observe texture dependent artifacts, that we believe to originate from a slight image blur in the cameras, which is due to a limited depth of field within our relatively large working volume, since the camera's operate with wide apertures set at $F = 2.0$. Compensation for this depth dependent Point Spread Function (PSF) by means of image deconvolution should remove these artifacts.

Acknowledgment

This work has been supported by the Integrated Graduate Program on Human-Centric Communication at Technische Universität Berlin.

References

[1] M. Waschbüsch, S. Würmlin, D. Cotting, F. Sadlo, and M. Gross, "Scalable 3d video of dynamic scenes," *The Visual Computer*, vol. 21, no. 8, pp. 629–638, 2005.

[2] A. Smolic, H. Kimata, and A. Vetro, "Development of mpeg standards for 3d and free viewpoint video," *Three-Dimensional TV, Video, and Display IV*, vol. 6016, 2005.

[3] F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, vol. 13, no. 1, p. 231, 2004.

[4] E. Stoykova, A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar *et al.*, "3-d time-varying scene capture technologies – a survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1568–1586, 2007. [Online]. Available: http://www.ics.forth.gr/ zabulis/B2.pdf

[5] A. Hosni, M. Bleyer, C. Rhemann, M. Gelautz, and C. Rother, "Real-time local stereo matching using guided image filtering," in *ICME, Workshop on Hot Topics in 3D Multimedia*, 2011.

[6] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3017–3024.

[7] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 321–334, 2004.

[8] M. Mueller, F. Zilly, C. Riechert, and P. Kauff, "Spatio-temporal consistent depth maps from multi-view video," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, may 2011, pp. 1 –4.

[9] M. Brown, D. Burschka, and G. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 993–1008, 2003.

[10] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1, pp. 7–42, 2002.

[11] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 519–528.

[12] J. Posdamer and M. Altschuler, "Surface measurement by space-encoded projected beam systems," *Computer graphics and image processing*, vol. 18, no. 1, pp. 1–17, 1982.

[13] J. Salvi, J. Pagès, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, pp. 827–849, 2004.

[14] S. Zhang and P. Huang, "High-resolution, real-time 3d shape acquisition," *Computer Vision and Pattern Recognition Workshop, 2004*, pp. 28–28, 2004.

[15] T. Weise, B. Leibe, and L. Van Gool, "Fast 3d scanning with automatic motion compensation," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*, 2007, pp. 1–8.

[16] P. Wissmann, R. Schmitt, and F. Forster, "Fast and accurate 3d scanning using coded phase shifting and high speed pattern projection," in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIM-PVT), 2011 International Conference on*, may 2011, pp. 108 –115.

[17] K. Ide, S. Siering, and T. Sikora, "Automating multi-camera self-calibration," in *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE, 2010, pp. 1–6.

[18] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multicamera self-calibration for virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 14, no. 4, pp. 407–422, 2005, camera calibration.