

# Modulation transfer function of patch-based stereo systems

Ronny Klowsky  
TU Darmstadt

Arjan Kuijper  
Fraunhofer IGD

Michael Goesele  
TU Darmstadt

## Abstract

A widely used technique to recover a 3D surface from photographs is patch-based (multi-view) stereo reconstruction. Current methods are able to reproduce fine surface details, they are however limited by the sampling density and the patch size used for reconstruction. We show that there is a systematic error in the reconstruction depending on the details in the unknown surface (frequencies) and the reconstruction resolution. For this purpose we present a theoretical analysis of patch-based depth reconstruction. We prove that our model of the reconstruction process yields a linear system, allowing us to apply the transfer (or system) function concept. We derive the modulation transfer function theoretically and validate it experimentally on synthetic examples using rendered images as well as on photographs of a 3D test target. Our analysis proves that there is a significant but predictable amplitude loss in reconstructions of fine scale details. In a first experiment on real-world data we show how this can be compensated for within the limits of noise and reconstruction accuracy by an inverse transfer function in frequency space.

## 1. Introduction

Patch-based (multi-view) stereo reconstruction [1, 4, 6, 7, 8] is a widely used technique to recover a 3D surface from photographs. Current methods achieve remarkable accuracy and are able to capture even fine geometric details [17, 13]. Their ability to faithfully reconstruct details is obviously limited by two facts: the sampling density of the algorithm and the size of the patch used for reconstruction (both of these are typically coupled to the resolution of the input images). To give a concrete example: a planar surface modulated with fine scale detail will eventually be reconstructed as a plane as image resolution decreases and patch size increases. This is illustrated in Figure 1 for a 1D signal.

We are interested in the geometry reconstructed by a patch-based algorithm for details that are roughly at the scale of the patch size. As also illustrated in Figure 1, such details are *reconstructed with much lower amplitude* and can even be inverted, so that valleys are reconstructed as

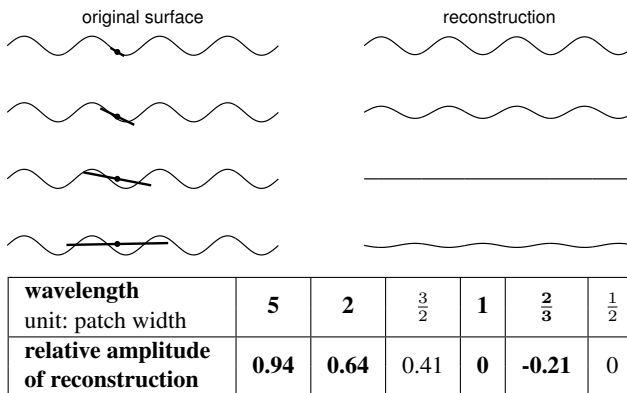


Figure 1. Predicted reconstruction of a sinusoidal surface with different patch widths. *Top*: The amplitude of the reconstruction varies drastically with the width of the patch used for reconstruction. In some cases, the signal is even inverted. The bold line marks the optimal patch position and orientation. *Bottom*: Table with predicted amplitude loss depending on patch width relative to signal wave length. Bold columns mark the cases drawn above.

peaks and vice versa. This behavior is not only contradicting our standard (or naïve) intuition about the properties of patch-based reconstruction, it is also in stark contrast to the assumptions made by most fusion techniques used to reconstruct a single surface from a set of reconstructed points or depth maps. These algorithms typically assume that the reconstructed points are samples of the true surface disturbed by zero-mean Gaussian noise [2, 9, 21]. Different scales or sampling densities are sometimes represented by lower confidences (or large variances in the noise model) and often enough just ignored. This implies that a reliable measurement of the true surface can be obtained by just averaging enough surface samples as this will cancel out noise.

In this paper, we show that there is a *systematic error* in the reconstruction depending on the details in the unknown surface (frequencies) and the reconstruction resolution. We show that even a “perfect” patch-based reconstruction algorithm will result in different reconstructed geometry of the same scene if used at different scales (e.g., varying resolution of input images or changing patch size). To our knowledge this fact is not modeled in any existing patch-based reconstruction algorithm. We provide a model that predicts

how amplitudes of different frequencies in the incoming signal are reproduced. The model is motivated by the concept of optical transfer functions (OTF) [19, 20] typically applied in the context of 2D image processing. It allows us theoretically to *invert this process*, in practice however only within the limits of noise and reconstruction accuracy.

The remainder of this paper is organized as follows: We first review related work (Section 2) before we derive and validate our model in 2D using synthetic examples and a real-world test target (Section 3). We then extend our theory to 3D (Section 4) and show its relevance on a real life application. Finally, we discuss our results (Section 5).

## 2. Related work

The analysis of different scale geometry reconstruction using patch-based stereo techniques has been neglected so far. For an overview and classification of multi-view stereo we refer to the recent survey and constantly updated benchmark by Seitz *et al.* [17, 13]. Key elements in our work build upon signal processing, optical transfer functions, and multi-scale surface representation. Existing work of the latter two areas will be discussed in the following.

The *optical transfer function* (OTF) is a well known concept to describe how details are reproduced by an imaging system [20]. It relies on the assumption of a linear system and describes how amplitude and phase change for different frequencies in the image using modulation and phase transfer functions, respectively. In our work, we validate that the linearity assumption holds and estimate the modulation transfer function of a patch-based stereo system. The OTF can be estimated in various ways [20]. For sampled imaging systems, Reichenbach *et al.* [16] introduced the knife-edge technique. Multiple scan lines are first registered to create a super-resolution edge profile and to suppress noise before the frequency space behavior is analyzed. Goesele *et al.* [5] applied this technique to estimate the modulation transfer function of a 3D range scanner. They capture a slanted edge and fit two planes to the measurements to create a super-resolution edge profile. The Fourier transform of the profile is then compared to that of an ideal edge.

Kobbelt *et al.* [10] define *multi-scale surface representations* and encode changes between levels using normal displacements. They use fairing operators to iteratively smooth a mesh and apply the results in the context of multi-scale surface editing. Inspired by Lindeberg’s scale-space theory [12], Pauly *et al.* [14] present a point-based multi-scale representation scheme using approximate geometric low-pass filtering and a projection operator to encode the different levels of detail. They discuss two approximate low-pass filters based on diffusion and least-squares filtering, respectively. Both can lead to deformations such as surface shrinkage. They identify the problem that no global, distortion-free parameterization exists for manifolds in general.

In this paper, we draw the *connection* between multi-scale surface representations and patch-based stereo reconstruction. We rely on the transfer function concept and the analysis techniques presented above, allowing us to demonstrate the effects in theory and practice. Using the simplifying assumption that the geometry can be represented as a height field, we are able to apply Fourier analysis to the reconstructed geometry.

## 3. Modeling the reconstruction process

The common strategy in patch-based stereo methods is to locally fit a planar patch to the unknown geometry that is photo-consistent with one or more other views. A typical example for measuring photo-consistency is the normalized cross-correlation (NCC) of points on the patch projected in other views. The final surface is represented by the (triangulated) central patch points [1, 4, 7] or the points are merged into a distance field [2, 3, 21]. In the following, we will develop a theoretical model for fitting a planar patch to the geometry, first in 2D and later in 3D (Section 4).

We assume that the geometry can be described as a height field  $z = f(x)$  (i.e., the whole surface is visible from an orthographic camera aimed perpendicular to the height field plane). In order to obtain the reconstruction  $\hat{z} = \hat{f}(x)$  at position  $x^*$  we fit a patch (line segment) with an extent of  $2\delta$  centered around  $x^*$  to the geometry. We represent the line segment by two parameters  $m, n$  and model the fitting process as optimizing for least-squares distance to the true geometry by minimizing the following energy

$$E(m, n, x^*) = \int_{x^*-\delta}^{x^*+\delta} (mx + n - f(x))^2 dx. \quad (1)$$

The reconstructed surface height at  $x^*$  is then given through the optimal parameters  $m, n$  by  $\hat{z} = mx^* + n$ . Note that we measure the patch extent along the  $x$ -axis in world coordinates and not in pixels as typically done in stereo. In the remainder of the paper we will use the term *patch width* for describing a patch of extent  $2\delta$ . The parameter  $\delta$  also depends on image resolution, surface distance to the camera, and the camera’s focal length. The actual patch size depends however on the slope (or orientation) of the patch. Intuitively, a smaller  $\delta$  allows to capture fine details whereas a larger  $\delta$  yields a smoothed surface. Image resolution often defines the sampling frequency equal to the distance between two consecutive points  $x_1^*$  and  $x_2^*$  where we fit a patch. In the following, we will deliberately disregard image resolution and think of reconstructing the geometry as fitting a patch continuously at every point  $x^*$ .

### 3.1. Theoretical results for a sine wave

We start by analyzing the simplest geometry in the sense of frequency behavior, a sine wave  $f(x) = a \sin(\omega x)$

with amplitude  $a$  and frequency  $\omega$ . To determine the reconstructed signal according to our model, we need to minimize  $E$  by finding the roots of the partial derivatives

$$\partial_m E = 2 \int_{x^*-\delta}^{x^*+\delta} x(mx + n - a \sin(\omega x)) dx \stackrel{!}{=} 0 \quad (2)$$

$$\partial_n E = 2 \int_{x^*-\delta}^{x^*+\delta} (mx + n - a \sin(\omega x)) dx \stackrel{!}{=} 0. \quad (3)$$

Solving the equations for  $m$  and  $n$  results in

$$m = \frac{3a \cos(\omega x^*) (\sin(\omega \delta) - \omega \delta \cos(\omega \delta))}{\omega^2 \delta^3} \quad (4)$$

$$n = \frac{a \delta^2 \omega \sin(\omega x^*) \sin(\omega \delta)}{\omega^2 \delta^3} + \frac{3a x^* \cos(\omega x^*) (\omega \delta \cos(\omega \delta) - \sin(\omega \delta))}{\omega^2 \delta^3} \quad (5)$$

Inserting this in  $\hat{z} = mx^* + n$ , the reconstruction is

$$\hat{f}(x^*) = \frac{a \sin(\omega \delta) \sin(\omega x^*)}{\omega \delta} = a \operatorname{sinc}(\omega \delta) \sin(\omega x^*). \quad (6)$$

This is an interesting result because frequency and phase of the sine are preserved for arbitrary patch width and frequency; only the amplitude is scaled by  $\operatorname{sinc}(\omega \delta)$  confirming one part of our linear system assumption. Note that for certain combinations  $\omega \delta$  the signal can even be inverted so that valleys become peaks and vice versa. In the following we will corroborate this result experimentally.

### 3.2. Experimental results for a sine wave

We first validate our results on synthetic data sets, rendered using the PBRT system [15]. This has the advantage that registration is perfect and all observed effects are due to photo-consistency optimization alone. As test target, we create a mesh representing a sine wave in the  $x, y$ -plane with  $z(x, y) = a \sin(\omega x)$ . The mesh is observed by five perspective cameras: One central camera points orthogonal to the  $x, y$ -plane and the other cameras are equally distributed around it with  $15^\circ$  parallax. A random texture with structure on all scales is mapped onto the geometry. We render views of the geometry using a variety of image resolutions. For the highest resolution we also create a ground truth depth map. For reconstruction, we run a patch optimization taken from an existing multi-view stereo system [6, Sect. 6.2] using the central camera as reference view and the surrounding cameras as neighbor views. For each pixel in the central camera the optimization is initialized with a fronto-parallel patch at depth values associated with that pixel in the highest-resolution ground truth depth map. The optimized patch with highest confidence (based on NCC) determines the depth at the current pixel.

For data analysis, we fit the parameters amplitude  $\hat{a}$ , frequency  $\hat{\omega}$ , phase  $\hat{p}$  and offset  $\hat{o}$  of the sine function

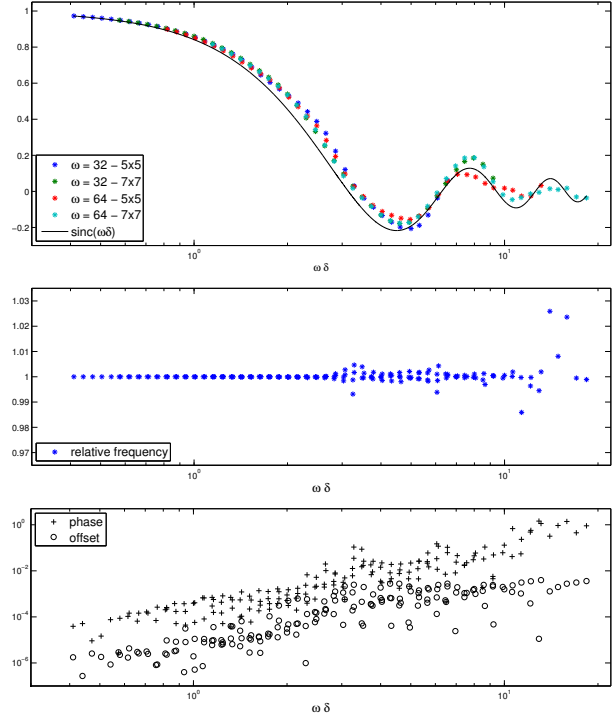


Figure 2. Resulting relative amplitude, relative frequency, phase, and offset of the reconstructed sine wave for different wavelengths and patch widths.

$z = \hat{a} \sin(\hat{\omega} x + \hat{p}) + \hat{o}$  to all reconstructed 3D points using Levenberg-Marquardt optimization [11]. To obtain a super-resolution sampling of the sine wave along the  $x$ -axis the camera’s up-vector is slightly tilted against the  $y$ -axis (about  $5^\circ$ ) similar to the knife edge technique [16]. In our experiments we use two sine waves of different frequency ( $\omega = 32$  and  $\omega = 64$ ). We vary the patch width parameter  $\delta$  by using various image resolution as well as image patch sizes of  $5 \times 5$  and  $7 \times 7$  pixels. Figure 2 shows that the reconstructed relative amplitudes, relative frequencies, phases, and offsets match very well with the predicted values. The observed differences are primarily caused by imperfections in the reconstruction process, in particular the interaction between the model texture and the photo-consistency of the patch.

### 3.3. Stereo transfer function

Ideally, we can express the reconstruction process using a transfer (or system) function representing the relation between input and output in terms of spatial frequencies. This concept is common in the imaging domain (optical transfer function) [19, 20] for describing the capability of showing fine details and the trade-off between blurred structure and aliasing. The optical transfer function is actually the Fourier transform of the point spread function. However, the transfer function concept is only applicable to linear systems featuring the principle of superposition and stationarity. The

latter is given for our model since the reconstruction is lateral shift invariant. What remains to check is the principle of superposition or additivity. We show that if the geometry is the sum of different frequency components the reconstruction is the sum of its separate contributions. For this purpose we represent  $f$  by a complete Fourier series

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)). \quad (7)$$

Substitution to Equation 1 and carrying out the same steps as in Section 3.1 one obtains

$$\hat{f}_\delta(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \text{sinc}(k\delta) \cdot (a_k \cos(kx) + b_k \sin(kx)) \quad (8)$$

as the reconstructed geometry (see Appendix A for intermediate steps). Thus, the principle of superposition is fulfilled and our model of patch-based stereo reconstruction is a linear system. This allows us to formulate the relationship between reconstructed and real geometry as

$$\hat{F}_\delta(\omega) = \text{MTF}_\delta(\omega) \cdot F(\omega) = \text{sinc}(\omega\delta) \cdot F(\omega) \quad (9)$$

where  $\hat{F}_\delta$  and  $F$  are the Fourier transforms of the reconstructed (using patch width  $2\delta$ ) and real geometry.  $\text{MTF}_\delta(\omega)$  is the modulation transfer function. Note that there is a difference to the traditional OTF. In our case the MTF can also be negative, modeling an inversion of amplitudes and the geometry, respectively. This allows us to completely remove the phase transfer function. In the next section, we will validate this result experimentally.

### 3.4. Experiments on a slanted edge

To experimentally validate Equation 9 we reconstruct a zigzag shape whose Fourier transform contains frequencies on all scales due to its sharp edges. Apart from the underlying geometry which is a zigzag shape (constant along  $y$ -axis) with edges of about  $126^\circ$  we use the same setup as in Section 3.2. Again, we just look at the  $(x, z)$  pairs of all reconstructed points. The slanted edge (implemented by the slightly tilted up-vector) gives us a fine sampling of the edge along the  $x$ -axis. We chose an interval  $[x_{min}, x_{max}]$  such that it captures exactly one period of the zigzag shape and sample all points therein into  $2^n$  bins so that the Discrete Fourier Transform (DFT) can be applied. In the Fourier transform of the ground truth profile every second coefficient is zero so we only use every second coefficient to compute the MTF, where the different resolutions lead to various patch widths  $2\delta$  (see Figure 3 top, middle). We can also measure how the amplitude is altered according to the product of frequency  $\omega$  and  $\delta$  (see Figure 3 bottom). Up to  $\omega\delta \approx 1.5$  the measured data matches very well with the theoretically predicted result. Beyond that point, the MTF still follows the theoretical prediction  $\text{sinc}(\omega\delta)$  but is masked by noise introduced by the reconstruction process.

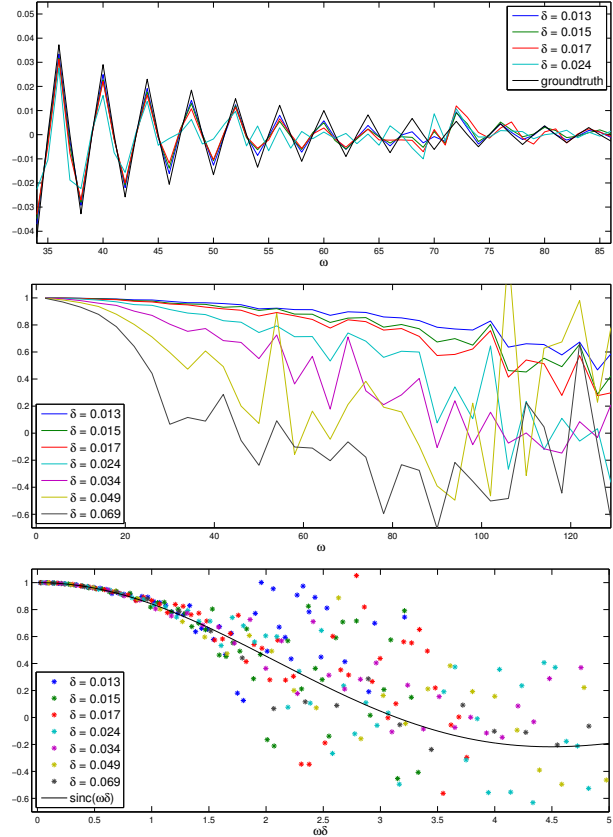


Figure 3. *Top*: Imaginary part of DFT coefficients for the zigzag profile. *Middle*: MTF samples for different patch widths  $2\delta$  as a function of  $\omega$ . *Bottom*: MTF as a function of the product  $\omega\delta$ .

### 3.5. Results on real-world data

Our goal is to analyze an object of simple and known 1D geometry to validate our theory with real world data. We therefore created a test target using 3D printing technology (see Figure 4). It consists of two periods of a sine wave with wavelength 62.8 mm and amplitude 10.0 mm and an edge with an angle of about  $126^\circ$ . Both are spread over 188.5 mm in width. To provide structure, we mapped the same texture as used in our synthetic experiments on the entire surface. This model was printed using a ZPrinter® 650 which has a printing accuracy of about 0.1 mm according to manufacturer specifications. For our experiments, we took photos with a digital SLR (one central photo looking orthogonal onto the object and several surrounding photos) with three different average camera distances to the object (near: 95 cm, middle: 145 cm, far: 280 cm). For each set of photos we perform a calibration using structure-from-motion [18]. We then apply a multi-view stereo algorithm [6] with patch-based optimization to compute a depth value for each pixel in the central views. Hereby, we repetitively rescale the images in order to get depth maps of different resolutions and



Figure 4. *Left*: Rendering of the test target. *Middle/Right*: Side and top view of the manufactured test target.

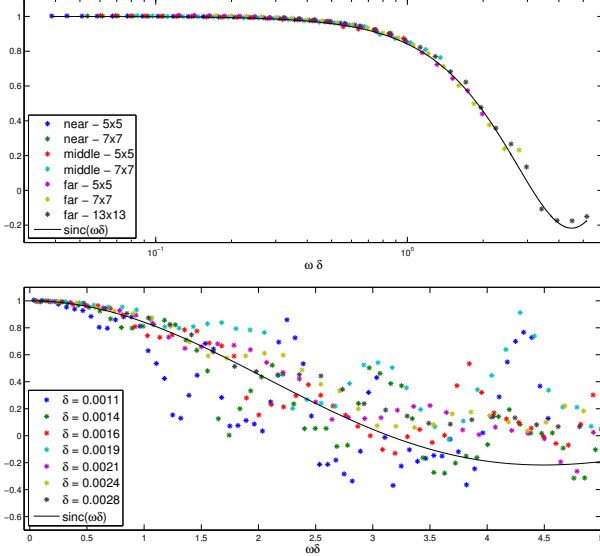


Figure 5. Results using the manufactured test target. *Top*: Amplitude loss on the sine wave. *Bottom*: MTF samples using the edge.

additionally run the reconstruction algorithm with two different image patch sizes ( $5 \times 5$  and  $7 \times 7$  pixel).

To analyze the amplitude loss on the sine wave, we first determine an optimal transform aligning the reconstruction with the  $x, y$ -plane. This optimal transformation is applied to all the different resolution depth maps to which we then fit in a second step a sine with amplitude, frequency, phase, and offset as in our synthetic experiments. Figure 5 (top) shows the amplitude loss with growing  $\omega\delta$ . The results closely match the theoretical prediction. In the second experiment, we analyze the reconstructed edge of the test target using an approach very similar to Goesele *et al.* [5]. We first fit two least-squares planes to the (highest resolution) reconstructed points on both sides of the edge and rotate the scan such that the intersection line coincides with the  $y$ -axis and the edge profile is symmetric to the  $y, z$ -plane. We then bin the reconstructed points ( $(x, z)$ -pairs) into 257 bins along the  $x$ -axis, move the ends to  $z = 0$  and multiply with a Blackman window. Then each profile is rotated around one end point to continue it periodically, dropping the first and last bin and thus resulting in 512 bins. We apply the Fourier transform to each profile and compare it to the Fourier transform of a perfect edge profile. Figure 5 (bot-

tom) shows the sampled MTF values for different  $\delta$ . The result shows significantly more noise and outliers than on the synthetic data reflecting errors in the registration, wrongly matched patches due to far-off start points and summed up errors during region growing.

## 4. Moving from 1D to 2D functions

So far, we described the theory for one-dimensional functions and validated it using geometry that is constant in one dimension. Naturally, real-world geometry rarely conforms to such a constrained model. We therefore show how our theory extends to height fields parameterized over a 2D plane, i.e., surfaces that can be described by  $z = f(x, y)$ .

### 4.1. Theory for a height field over a 2D plane

Similar to the one-dimensional case we fit a planar patch  $P = m_x x + m_y y + n$  to the height field around a point  $(x^*, y^*)$  with dimensions  $\delta \times \delta$ . We again express the geometry in terms of a sine and cosine series or, for convenience, using complex numbers

$$f(x, y) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_{j,k} e^{i(jx+ky)}. \quad (10)$$

Solving the according minimization problem

$$\min_{m_x, m_y, n} \int_{y^*-\delta}^{y^*+\delta} \int_{x^*-\delta}^{x^*+\delta} (P - f(x, y))^2 dx dy \quad (11)$$

for  $m_x, m_y$ , and  $n$  we obtain the reconstructed geometry as

$$\sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \text{sinc}(j\delta) \text{sinc}(k\delta) \quad (12)$$

(see Appendix C for a thorough derivation handling the more general case of a rectangular patch). Note that this amplitude loss is a product of two sinc functions which is the Fourier transform of a box filter.

### 4.2. Results on synthetic 2D sine

We will substantiate the theoretical result on geometry containing only one frequency along each dimension and construct a height field with  $z = \frac{1}{\omega} \sin(\omega x) \sin(\omega y)$ . Apart from this geometry, the setup is equivalent to that in Sections 3.2 and 3.4. We optimize for the six parameters amplitude  $\hat{a}$ , frequencies  $\hat{\omega}_x, \hat{\omega}_y$ , phases  $\hat{p}_x, \hat{p}_y$ , and offset  $\hat{o}$  such that  $z = \hat{a} \sin(\hat{\omega}_x x + \hat{p}_x) \sin(\hat{\omega}_y y + \hat{p}_y) + \hat{o}$  holds for the reconstructed 3D points. According to the theoretical result from Equation 12, the reconstructed amplitude should be scaled by  $\text{sinc}^2(\omega\delta)$  compared to the original amplitude. Figure 6 shows that the experimentally obtained scaling factors match the expected values very well. The estimated frequencies, phase shifts, and offsets are comparable to the 1D experiments (similar to Figure 2).

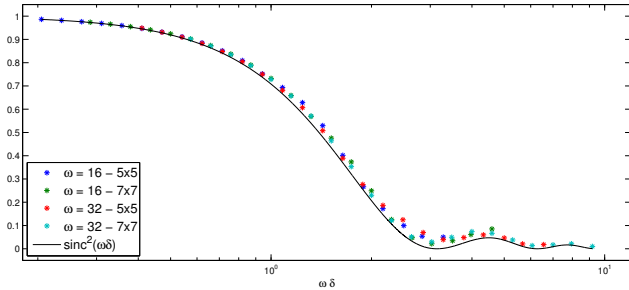


Figure 6. Reconstructed amplitude as fraction of the true amplitude compared to theoretical prediction in 2D.



Figure 7. *Left*: Sample image of the lion head sculpture. *Right*: Low-resolution VRIP reconstruction.

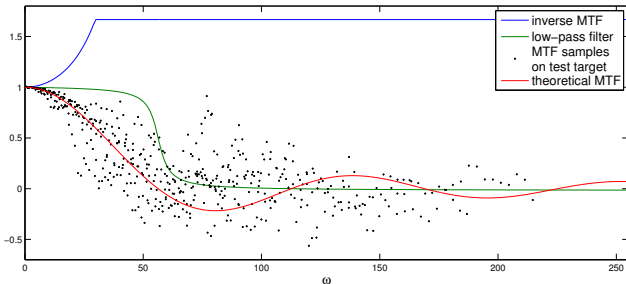


Figure 8. Slice of the 2D inverse MTF and of the low-pass filter used for the lion head experiment.

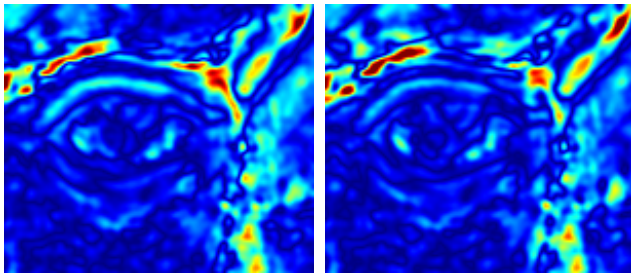


Figure 11. Absolute depth differences of results shown in Figure 10 left/middle compared to Figure 10 right. Note the changes around the eyelid and the nose.

### 4.3. Application to real-world example

After presenting all the theoretical results and experiments validating the results in practice, we want to exploit the new insights within a real-world application. In the fol-

lowing we enhance a single-scale multi-view stereo reconstruction. For that purpose we create a 3D model of a lion head sculpture using the following pipeline. We register 225 photographs [18] of a lion head sculpture, reconstruct a depth map for a subsets of 41 views with image patch size of  $7 \times 7$  pixels [6], and merge the depth maps into a global model using VRIP [2] (see Figure 7). Hereby, we create two different versions, a low-resolution model using downsampled photos (halved image dimensions) for depth map reconstruction and a high-resolution model using full image resolution. We convert a cut-out of the models into a height field and smoothly interpolate to a constant value and zero gradient at the borders minimizing second order derivatives. This leads to a periodical signal which is the input to a 2D Fourier transform. For all frequencies, we compute the inverse MTF using our model and scale up the frequencies accordingly to invert the amplitude loss during reconstruction. Since our experiments showed significant noise and thus deviation from the ideal MTF for the real-world test target, we clamp the inverse MTF. We use  $MTF_{\delta}(\omega)^{-1} = \min(0.6^{-1}, \text{sinc}(\omega\delta)^{-1})$  (Figure 8). We also apply a smooth low-pass filter that suppresses high-frequencies where the patch size is smaller than the wavelength. Finally, the inverse Fourier transform is performed. Figures 9 and 10 show how details are emphasized through the inversion of our stereo transfer function. Difference images in Figure 11 show a quantitative comparison where some regions are improved whereas others become worse.

## 5. Discussion

We introduce a theoretical model of patch-based stereo, modeling the reconstruction process as a linear system, and validate it on synthetic and real input using an existing multi-view stereo system. We demonstrate that there is a *significant amplitude loss* and even an inversion of amplitudes which has *not been modeled before* in any of the existing reconstruction pipelines. The real-world application example gives a first clue of how this could improve the reconstruction quality in a practical system. Inevitably, the experiments show some limitations. First, modeling the reconstruction process as finding the depth and orientation of a patch that minimizes the least-squares distance to the true surface leaves out the complex interaction between the surface texture and the reconstruction. This may yield artifacts when the MTF is inverted. Second, the noise introduced in the reconstructions may of course limit the ability to invert the amplitude loss. Finally, practical applicability is limited because of the nature of the global Fourier method causing problems with depth discontinuities (occlusion), finite image size (periodicity assumption), and incomplete reconstructions. The lion head example is therefore only a starting point of how geometry can be reconstructed faithfully using our amplitude loss compensation.

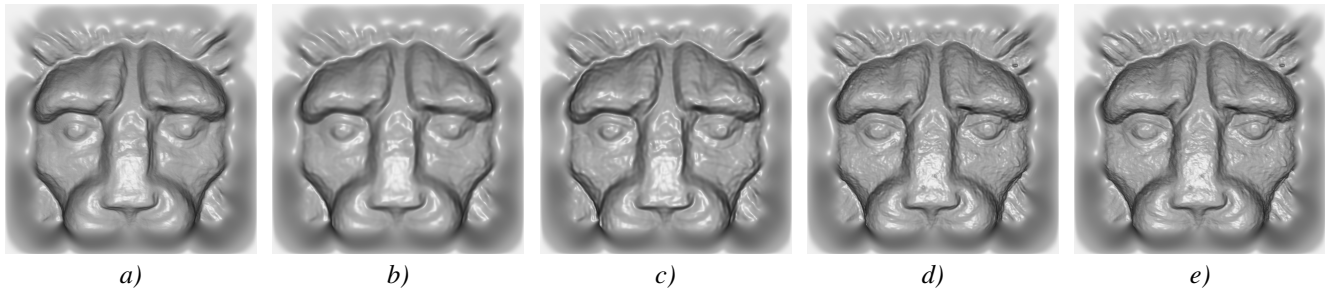


Figure 9. Results on a height field created from the lion head VRIP model. *a)* Low-resolution reconstruction. *b)* Removed high-frequency noise. *c)* Inverted amplitude loss up to a certain scale. *d)* Smoothed high-resolution reconstruction. *e)* High-resolution reconstruction.

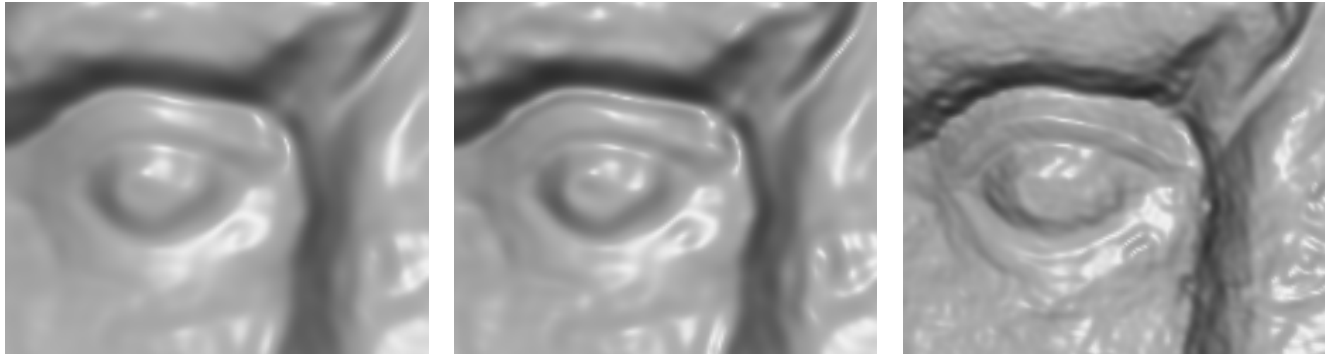


Figure 10. *From left to right:* Magnification of a region around the left eye in Figure 9 *b)*, *c)*, and *d)*, clearly showing how our proposed method improves the details, e.g., of the eyelid. See Figure 11 for a quantitative visualization of the differences.

**Acknowledgements** This work was supported in part by the DFG Emmy Noether fellowship GO 1752/3-1.

## References

- [1] D. Bradley, T. Boubekeur, and W. Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *CVPR*, 2008. 1, 2
- [2] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH*, 1996. 1, 2, 6
- [3] S. Fuhrmann and M. Goesele. Fusion of depth maps with multiple scales. In *SIGGRAPH Asia*, 2011. 2
- [4] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *PAMI*, August 2010. 1, 2
- [5] M. Goesele, C. Fuchs, and H.-P. Seidel. Accuracy of 3D range scanners by measurement of the slanted edge modulation transfer function. In *3DIM*, 2003. 2, 5
- [6] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *ICCV*, 2007. 1, 3, 4, 6
- [7] M. Habbecke and L. Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In *CVPR*, 2007. 1, 2
- [8] M. Jancosek, A. Shekhovtsov, and T. Pajdla. Scalable multi-view stereo. In *3DIM*, 2009. 1
- [9] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *EG SGP*, 2006. 1
- [10] L. Kobbelt, S. Campagna, J. Vorsatz, and H.-P. Seidel. Interactive multi-resolution modeling on arbitrary meshes. In *SIGGRAPH*, 1998. 2
- [11] Levenberg-Marquardt nonlinear least squares algorithms in C/C++. <http://www.ics.forth.gr/~lourakis/levmar/>. 3
- [12] T. Lindeberg. *Scale-Space Theory In Computer Vision*. Kluwer Academic Publishers, 1994. 2
- [13] Middlebury multi-view stereo benchmark. <http://vision.middlebury.edu/mview/>. 1, 2
- [14] M. Pauly, L. P. Kobbelt, and M. Gross. Point-based multi-scale surface representation. In *ACM TOG*, 2006. 2
- [15] Physically based rendering. <http://www.pbrt.org>. 3
- [16] S. E. Reichenbach, S. K. Park, and R. Narayanswamy. Characterizing digital image acquisition devices. In *Optical Engineering*, 1991. 2, 3
- [17] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, 2006. 1, 2
- [18] N. Snavely, S. M. Seitz, and R. Szeliski. Skeletal sets for efficient structure from motion. In *CVPR*, 2008. 4, 6
- [19] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, New York, 2010. 2, 3
- [20] T. L. Williams. *The Optical Transfer Function of Imaging Systems*. Institute of Physics Publishing, 1999. 2, 3
- [21] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust TV-L1 range image integration. In *ICCV*, 2007. 1, 2

## A. Geometry reconstruction in 1D

Representing the underlying geometry  $f(x)$  with a complete sine and cosine series yields

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)). \quad (13)$$

We need to find  $m$  and  $n$ , so that on the interval  $I = [x^* - \delta, x^* + \delta]$  the energy  $E(m, n, x^*)$  is minimized:

$$E(m, n, x^*) = \int_I (mx + n - f(x))^2 dx. \quad (14)$$

This implies taking partial derivatives with respect to  $m$  and  $n$  and finding the roots of these equations:

$$\begin{aligned} \partial_m E(m, n, x^*) &= \int_I 2x (mx + n - f(x)) dx \stackrel{!}{=} 0 \\ \partial_n E(m, n, x^*) &= \int_I 2 (mx + n - f(x)) dx \stackrel{!}{=} 0. \end{aligned} \quad (15)$$

This yields the following solution for Equation 15:

$$\begin{aligned} E_m &= nx^2 + \frac{2}{3}mx^3 - \frac{1}{2}a_0x^2 + \sum_{k=1}^{\infty} \frac{2}{k^2} (-a_k \cos(kx) \\ &\quad - b_k \sin(kx) - ky a_k \sin(kx) + ky b_k \cos(kx)) \\ E_n &= 2nx + mx^2 - xa_0 \\ &\quad - \sum_{k=1}^{\infty} \frac{2}{k} (a_k \sin(kx) - b_k \cos(kx)). \end{aligned} \quad (16)$$

Inserting the boundaries of the interval  $I$  (ignoring the superscript  $*$  for typographic reasons) in Equation 16 yields

$$\begin{aligned} 0 &= 4nx\delta + 4mx^2\delta + \frac{4}{3}m\delta^3 - 2x\delta a_0 + \sum_{k=1}^{\infty} \frac{4}{k^2} ( \\ &\quad - xka_k \cos(kx) \sin(k\delta) - \delta ka_k \sin(kx) \cos(k\delta) \\ &\quad + a_k \sin(kx) \sin(k\delta) - xkb_k \sin(kx) \sin(k\delta) \\ &\quad + \delta kb_k \cos(kx) \cos(k\delta) - b_k \cos(kx) \sin(k\delta)) \\ 0 &= (-4\delta mx - 4\delta n + 2\delta a_0 + \\ &\quad \sum_{k=1}^{\infty} \frac{4}{k} \sin(k\delta) (a_k \cos(kx) + b_k \sin(kx))) \end{aligned} \quad (17)$$

These two equations are linear in  $m$  and  $n$  and can be easily solved. Moreover, from Equation 17 one obtains the expression for the solution  $mx + n$  directly as

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \text{sinc}(k\delta) (a_k \cos(kx) + b_k \sin(kx)). \quad (18)$$

## B. 1D reconstruction with complex numbers

Using complex numbers, we can write our series as

$$f(x) = \sum_{k=0}^{\infty} \alpha_k e^{ikx} \quad (19)$$

and solve the expressions for the partial derivatives as the following indefinite integrals

$$\begin{aligned} E_m &= nx^2 + \frac{2}{3}mx^3 + 2 \sum_{k=0}^{\infty} \alpha_k e^{ikx} \left( \frac{-1}{k^2} + \frac{ix}{k} \right) \\ E_n &= 2nx + 2mx^2 + \sum_{k=0}^{\infty} \frac{2i}{k} \alpha_k e^{ikx}. \end{aligned} \quad (20)$$

Restricting to the interval  $I$  we get

$$\begin{aligned} E_m &= 4nx\delta + 4mx^2\delta + \frac{4}{3}m\delta^3 + \\ &\quad \sum_{k=0}^{\infty} \alpha_k e^{ikx} \frac{4i}{k^2} (k\delta \cos(k\delta) + (-1 + ikx) \sin(k\delta)) \\ E_n &= 4\delta(n + mx) - \sum_{k=0}^{\infty} \frac{4}{k} \alpha_k e^{ikx} \sin(k\delta). \end{aligned} \quad (21)$$

From this,  $m$  and  $n$  can be solved resulting in

$$mx + n = \sum_{k=0}^{\infty} \text{sinc}(k\delta) \alpha_k e^{ikx} \quad (22)$$

## C. 2D reconstruction with complex numbers

Clearly, the same procedure can be applied in 2D. Let

$$P = m_x x + m_y y + n \quad (23)$$

be the solution to the patch that we want to compute around a point  $(x, y)$  spanned by  $I = [x - \delta, x + \delta] \times [y - \epsilon, y + \epsilon]$ . The signal  $f(x, y)$  can be expressed in terms of a sine and cosine series or, alternatively, using complex numbers by

$$f(x, y) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_{j,k} e^{i(jx+ky)} \quad (24)$$

Again we want to find the minimum of

$$E = \int_y \int_x (P - f)^2 dx dy \quad (25)$$

for the parameters  $m_x$ ,  $m_y$ , and  $n$ . Taking derivatives with respect to these parameters and solving yields

$$\begin{aligned} E_{m_x} &= nx^2 y + \frac{2}{3}m_x x^3 y + \frac{1}{2}m_y x^2 y^2 \\ &\quad + \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \left( \frac{2i}{j^2 k} + \frac{2x}{jk} \right) \end{aligned} \quad (26)$$

$$\begin{aligned} E_{m_y} &= nxy^2 + \frac{1}{2}m_x x^2 y^2 + \frac{2}{3}m_y xy^3 \\ &\quad + \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \left( \frac{2i}{j^2 k} + \frac{2y}{jk} \right) \end{aligned} \quad (27)$$

$$\begin{aligned} E_n &= 2nxy + m_x x^2 y + m_y xy^2 \\ &\quad + \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \frac{2}{jk}. \end{aligned} \quad (28)$$

On the given patch  $I$  we get

$$\begin{aligned} E_{m_x} &= nx + m_x x^2 + m_y xy + \frac{1}{3}m_x \delta^2 \\ &\quad + \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \sin(k\epsilon) \cdot \\ &\quad \left( \frac{i}{jk\epsilon} \cos(j\delta) - \sin(j\delta) \left( \frac{i}{j^2 k \delta \epsilon} + \frac{x}{jk \delta \epsilon} \right) \right) \end{aligned} \quad (29)$$

$$\begin{aligned} E_{m_y} &= ny + m_x xy + m_y y^2 + \frac{1}{3}m_y \epsilon^2 \\ &\quad + \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \sin(j\delta) \cdot \\ &\quad \left( \frac{i}{jk\delta} \cos(k\epsilon) - \sin(k\epsilon) \left( \frac{i}{j k^2 \delta \epsilon} + \frac{y}{jk \delta \epsilon} \right) \right) \end{aligned} \quad (30)$$

$$\begin{aligned} E_n &= n + m_x x + m_y y \\ &\quad - \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \frac{1}{jk \delta \epsilon} \sin(j\delta) \sin(k\epsilon) \end{aligned} \quad (31)$$

We can solve these linear equations in  $m_x$ ,  $m_y$ , and  $n$ . From  $E_n = 0$  one can directly derive the solution for our patch:

$$\begin{aligned} P &= m_x x + m_y y + n \\ &= \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \frac{1}{jk \delta \epsilon} \sin(j\delta) \sin(k\epsilon) \\ &= \sum_{j,k} \alpha_{j,k} e^{i(jx+ky)} \text{sinc}(j\delta) \text{sinc}(k\epsilon). \end{aligned} \quad (32)$$