Cultural Heritage

High resolution acquisition of detailed surfaces with lens-shifted structured light

M. Ritz a,*, F. Langguth b, M. Scholz b, M. Goesele b, A. Stork a,b

Abstract

We present a novel 3D geometry acquisition technique at high resolution based on structured light reconstruction with a low-cost projector–camera system. Using a 1D mechanical lens-shifter extension in the projector light path, the projected pattern is shifted in subpixel scale steps with a granularity of up to 2048 steps per projected pixel, which opens up novel possibilities in depth accuracy and smoothness for the acquired geometry. Combining the mechanical lens-shifter extension with a multiple phase shifting technique yields a measuring range of 120 × 80 mm while at the same time providing a high depth resolution of better than 100 μm. Reaching beyond depth resolutions achieved by conventional structured light scanning approaches with projector–camera systems, depth layering effects inherent to conventional techniques are fully avoided. Relying on low-cost consumer products only, we reach an area resolution of down to 55 μm (limited by the camera). We see two main benefits. First, our acquisition setup can reconstruct finest details of small cultural heritage objects such as antique coins and thus digitally preserve them in appropriate precision. Second, our accurate height fields are a viable input to physically based rendering in combination with measured material BRDFs to reproduce compelling spatially varying, material-specific effects.

1. Introduction

Digitizing objects in three dimensions opens up a rich set of possibilities for appropriate and long-term preservation, especially in the case of cultural heritage objects, as opposed to digitization only as 2D images. Since the human vision system relies on visual depth cues, 3D models can reveal a far more immersive and tangible impression of the object’s features and significance due to the infinite possibilities of choosing viewpoint, lighting, and rendering settings at any time after acquisition, as opposed to the rather fixed nature of 2D images. 3D scanning with structured light is, however, limited, and does not resolve fine details with sufficient quality. We present and analyze a 3D scanning approach that has a high depth resolution, allowing for the reconstruction of finest surface details. Given the fine marble structure of sculptures or structures found on very small and movable objects such as coins, capturing of finest details can be crucial to appropriate long-term preservation in this object domain. Besides digitization, analysis of aging effects such as micro-cracks in objects is in the interest of museums, as well as the reconstruction of fine engravings, which current techniques have not yet mastered. Applications reach even into the domain of art where determining the depth of paint can reveal specific information on the artist. Material acquisition at fine-granular scale is also a vital necessity in the field of realistic 3D rendering, where the focus changes from using approximate close-to-reality shading models to rendering physically acquired reflections of materials, due to the increasingly powerful hardware delivering compelling realism. Some essential information is the geometry of the material samples, that provides the basis necessary for realistic rendering of spatially varying bidirectional reflectance distribution functions (SVBRDFs), or provide rendering systems with geometry information for physically correct light throughput regarding traversal of fine material structures.

Based on the previously published version of this paper [1], we propose a fully automatic geometry acquisition system with the following contributions:

1. novel lens-shifting reconstruction technique based on a precise hardware shifting extension [2], combined with multiple phase shifting [3], lifting common structured light limitations,
2. accuracy analysis of the system,
3. high resolution capture at comparatively large measuring range and low cost due to off-the-shelf components,
4. accurate height field acquisition for material-realistic rendering,
5. qualitative and quantitative comparison against a commercial 3D scanner with comparable specifications.
• evaluation of optimal lens-shift image count/shifting step size, and
• introduction of a model-intrinsic per-sample confidence metric.

2. Related work

Various methods address 3D digitization [4], including structured light and time of flight scanning, as well as photometric methods such as shape from specularity or shape from shading. Photometric methods are often based on specific material properties. For example, specular reflections are used to derive a normal field by moving a light source over the object [5]. Francken et al. [6] project several binary patterns on the object using an LCD-display to obtain a unique identification of the incoming light direction at maximum intensity for surface normal computation. A disadvantage of the above techniques is their dependence on specular reflectance, whereas an advantage over other methods is the ability to measure the surface of translucent objects.

Glencross et al. [7] developed an approach that can acquire meso-structure using only two images of an object. Although having the advantage of a very minimal capturing setup this technique only produces a visually realistic surface model which is optimized for rendering and not evaluated in terms of high geometric accuracy. For improved precision some work also combines photometric and structured light methods [8].

Our approach is based on pure structured light which has been widely investigated and has some significance in the industry. Techniques in this field include projector–camera systems. Several cameras can be used to increase robustness and accuracy, often leading to a very complex capturing setup [9]. Acquisition is done by projecting a series of patterns onto the object to obtain a unique correspondence between projector-planes and camera rays to determine 3D points by intersection. The basic approach uses gray code patterns to minimize the effect of 1-bit measuring errors by a special encoding. The sequence in which a pixel is overlaid by dark and bright stripes reveals the plane ID. To overcome several short-range effects such as defocus more advanced patterns were studied by Kim et al. [10]. Another method extending mainly the patterns projected by using color coding is applied by [11]. The downside of these methods is a layering effect in depth, since all surface positions lying on a pattern stripe are clustered together, leading to a discretization in space during intersection. This can be improved by multiple phase shifting [3], where three different stripe patterns formed by different wavelengths of a sinusoidal luminance function are used instead of binary patterns. Each pattern is then shifted by fractions of its wavelength over the object, revealing a certain phase for each surface point. All patterns combined allow for globally unique determination of the correlation between surface points and projector plane IDs. Shifting the patterns removes the layering effect somewhat as the discretization is more fine granular, by allowing the determination of stripe IDs on a subpixel basis. Problems arise in general when objects show a high percentage of indirect light transport, as well as for dark materials and surfaces with high specularity.

Global illumination effects are tackled by Chen et al. [12] using modulated high- and low-frequency pattern projections. The observation that the integration of a normal field mostly yields low frequency errors, while structured light scanners are mainly susceptible to high frequency errors, leads to a combination of photometric methods with range scanning methods, as done by Nehab et al. [13]. They fuse both domains of surface descriptions using an optimization technique that increases the quality of 3D scans. Other approaches such as [14] use specific projection patterns, that are resilient to individual global illumination effects.

A lot of current research also aims for real time acquisition by using high speed capturing devices [15], specialized optimization techniques [16], as well as GPU-based implementations [17]. However, these techniques often suffer from increased accuracy losses.

Finally, going into the domain of line shifting scanners, a digital projector is used in [18] to project thin stripes in regular distances, which are then moved over the object. For each camera pixel, the maximum luminance over scanning time and the stack of all patterns is determined, enabling a precise correlation between projector and camera pixels. In comparison to phase shifting approaches, this technique is more robust against indirect light transport due to larger distances between projected stripes, but shows problems with specular reflections.

Even though the field of phase shifting has been widely explored, to our knowledge there are currently no publications closely related to our acquisition setup. We build on the multiple phase shifting technique but drastically increase its depth resolution by adding an additional pattern at the projector’s Nyquist limit. Since phase shifting is in this case no longer possible using different projected patterns, we use a hardware lens-shifter to shift the pattern with subpixel accuracy. As a result, we fully avoid depth layering effects inherent to structured light techniques and moreover break limits in depth accuracy, going far beyond 100 μm.

3. Hardware setup and analysis

The acquisition setup consists of a DLP LED projector, a consumer DSLR (digital single-lens reflex) camera, the lens-shifter and the sample mount (Fig. 1). These constituents are mounted rigidly into an alloy framework covered with a matte black anti-reflective foil to
minimize stray light. The framework itself rests on rubber pads to decouple it from external vibrations and to damp vibrations caused by the camera. This setup makes the system insensitive against minor external vibrations.

The angle between camera and projector is approximately 20°. The baseline measures about 170 mm. The average measurement distance to the sample mount is 250 mm. Together with the projector and camera intrinsic parameters, this leads to a measurement area of about 120 × 80 mm covered by 2176 × 1434 sample points. The optical parameters of the projector and camera currently used limits the depth of the scanning volume to about 3 cm, which is the range where both the projection and the camera image are in focus. Note that for the current setup, we obtained the above measurements manually as opposed to using an automatic calibration method, which will be a next step of extension.

3.1. Projector and camera

The projector used in the system is an LG-HS101 Pocket Projector with a resolution of 800 × 600 pixels using high performance LEDs in the three color channels. It is directly fed with image data by the reconstruction algorithm. As capturing device, a Nikon D300S was chosen with a resolution of 4352 × 2868 pixels (12 MPix). Its image sensor consists of periodic groups of 2 × 2 color sensors, 2 for green color, and one for red and blue, respectively. Rather than using the data postprocessed by the camera, which interpolates the color values for each of the four positions in the groups, we directly read out raw data from the camera and form a per-color weighted average manually from the exact readings per sensor unit to obtain a grayscale measurement for the pixel group. Since every physical pixel of the sensor chip can detect only one of the three primary colors, this procedure is necessary and leads to an effective resolution of 2176 × 1434. We use a Zeiss Distagon macro-objective (1:2/35) with low distortion and small color aberration. The camera is integrated into the system over a USB port, used both for fully automatic camera control and image data gathering. The camera as the only measuring component in the system is strongly influenced by image noise, which has a direct influence on the accuracy of reconstruction. Thus, the representative measure of signal to noise ratio (SNR) was optimized over the space of exposure settings in order to maximize reconstruction quality. The use of a specialized industrial camera rather than the DSLR (digital single lens reflex) currently used would increase the number of images captured per second and avoid vibrational influence. Increasing the resolution beyond 12 MPix would clearly improve acquisition accuracy, but at the same time lead to the logical consequence of using a higher-resolution projector. With increased resolution of both the camera and the projector, the total cost of the system rises significantly, which is why we are currently staying with the Nikon D300S.

3.2. Lens-shifter

The major contribution of our method is the use of a mechanical lens-shifter extension [2] and a lens-shifting reconstruction technique to significantly increase depth resolution. The lens-shifter is a brass frame with an embedded lens and four flexure joints at the corners, realized by very thin regions of the material. They allow for nearly friction-less 1 DOF movement of the frame and thus enable smooth and accurate shifting of the lens with precise repeatability. The lens-shifting extension is based on existing hardware. The motor used for moving the frame is a magnetic voice coil motor responsible for moving the heads of a standard hard disk drive. Together with the controller needed to provide the necessary current to drive the lens-shifter and to provide an interface for steering commands to the reconstruction software, the cost is still significantly lower than that of the projector or camera alone. The controller driving the shifter is integrated into the system over the serial port which is used to transmit steering commands.

The lens-shifter mechanism serves two purposes. Firstly, the lens itself improves the projector’s depth of focus and allows for a smaller distance between projector and object to increase reconstruction area resolution. Secondly, the projected image can be moved along one axis at subpixel accuracy over a range of ±1 projection pixel in a granularity of 2048 steps per pixel. The maximum number of shifting steps is defined by the hardware design of the controller which can drive an electric current in as many different intensity steps through the inductor of the lens shifter to create a proportional shifting force. With the precise shifting of the projection, high frequency patterns can be used for the scan process, which amplifies spatial differences in luminance measured for small surface profile differences. These two features make a more dense sampling of the projected signal possible, yielding a higher depth accuracy of the scanner. In addition, smaller pattern wavelengths increase the system’s robustness against indirect light transport as observed by [12].

The lens-shifter is responsible for precise shifting of the projection. Positioning errors have direct influence on the luminance measured by the camera and thus on the reconstruction, making repeatability a highly significant measure for consistent and accurate scans. To assess repeatability, a small mirror was mounted at the side of the

Fig. 2. Left: lens-shifter. Middle: drawing with color coded displacement of the frame under current-induced Lorenz force due to magnetic field. Right: laser light path (sketched) from source (upper right) over lens-shifter mirror and external mirror to projection surface. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
lens-shifter frame to redirect a laser beam onto the projection surface captured by the camera, and extend the light path length to register even minimal elongations of the shifting device, measurable as the movements of the laser dot seen by the camera (cf. Fig. 2). The shifter was repeatedly driven from the neutral position to predefined positions, and overall repeatability was found to be within an error of 0.001 projector pixels. The laser method also confirmed that the linearity between the desired position, conveyed by the current driven by the controller, and the actual position of the lens-shifter, is granted and sufficient.

4. Meso-structure acquisition

The process of capturing images with patterns overlaid over the object is done in two phases, one for capturing phase shifted images for three different wavelengths according to the method of Lilienblum et al. [3], and one for capturing lens-shifted images. The first phase assigns surface points to a certain period and resolves the uniqueness problem of the second phase, which is responsible for exact determination and ray-plane-intersection based on the lens-shifting information. The capturing phase is followed by normalizing all captured images (see Section 5.1) and the actual reconstruction of all surface points seen by the camera. Finally, a surface mesh is constructed between the samples of the reconstructed point cloud. Since each 3D point correlates with the 2D image point from which it was reconstructed, and all 2D image points lie in a regular grid, the connectivity information applies to the 3D point cloud as well. Triangulation can be done in a straightforward way by inserting two triangles between any rectangular neighborhood of four points.

4.1. Lens-shifting

The second step of the acquisition is also based on the phase shifting approach similar to [3], but with a single wavelength and finer steps. A sine wave pattern with fixed wavelength is projected and shifted by a fixed distance after each capturing. Instead of shifting the pattern digitally, the lens-shifter is used for this task, while the projector output remains unchanged. Since it can position the pattern independent of the projector resolution at subpixel accuracy with a theoretical step size of \( \frac{1}{2048} \) of a projector pixel, the use of high-frequency patterns is now possible. Thus, the pattern with the smallest displayable wavelength of two was used for reconstruction, consisting of one pixel wide black and white bars. Note that insufficiencies in the optical system convert this binary signal into a close approximation of a sinusoidal signal. The acquisition yields a luminance profile per pixel which we interpreted as a time-dependent signal.

There is a significant difference between digital and analogous shifting of the signal. Regular phase shifting produces the same luminance profile for each pixel due to the time dependent signal not being coupled to its spatial dimension. However, in the case of the optical lens-shifting where a coupling between time and space is given, the time dependent signal is impacted by lens distortions and other optical influences, and both temporal and spatial differences in the wavelength for different image positions are the consequence. The conventional phase shifting approach of using Fourier analysis for phase determination is thus not applicable because it requires the exact wavelength. Instead, an optimization problem is solved to fit the luminance profile into a model function per pixel. With the resulting accurate phase, projector and camera pixel can be correlated uniquely within the domain of a wavelength.

4.2. Fitting the model function

Despite steering the projector with a rectangular signal, the luminance pattern of each pixel over a lens shifter sequence resembles a sine wave. The reason is probably the optics and projector image generator that lead to a rather washed-out

---

*Fig. 3.* Upper left: Luminance for four distinct pixels over common full lens-shifter sequence (i.e. \( \pm 2048 \) steps as described in Section 3.2, horizontal axis). Luminance flow over full sequence before image normalization (upper right), after image normalization (lower left), and after 5px box filtering (lower right). Vertical axis: amplitude normalized to \( \pm 1 \). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
reproduction of the image, resulting in a sine wave shaped luminance pattern registered in each pixel over the shifting sequence. Determining the phase of the signal requires the model function to be periodic and to well represent the captured data.

The analysis of the normalized temporal signal of several distinct image positions is shown in Fig. 3 (upper left). Per-pixel normalization is necessary to account for inhomogeneous distribution of reflectance of most objects, especially textured surfaces, where the signal amplitude varies strongly over space. The following model function is used, parameterized in amplitude \(a\), frequency \(f\), phase \(\phi\), and offset \(d\).

\[
b(x) = a \sin(\phi + 2\pi f x) + d
\]  

(1)

We also need to fit the offset in order to deal with indirect lighting (see [19]). The four parameters are then found by optimization using the Levenberg–Marquardt algorithm for nonlinear least squares problems [20]. Fig. 4 shows an example of a fitting result. Appropriate starting values are necessary for a robust optimization (see Section 5). For the spatially varying phase, the value is left variable. The starting value for the amplitude is set to \(a = 1\) due to the normalization, the frequency is set empirically.

4.3. MPS and LS combined: high depth accuracy at large range

The relative and more accurate phase \(\phi\) determined by lens-shifting is now combined with the absolute and coarser phase \(\omega\) resulting from classical phase shifting. The globally unique phase \(\omega\) is transformed to the range [0,1], corresponding to a pixel column within the bounds of the projected image. \(\phi\) is a local position within the wavelength \(\lambda\) and has the same range as \(\omega\), but in this case mapping to a position within the bounds of a lens shifting wavelength:

\[
x = \lambda \left( \frac{\omega}{\lambda} - \phi + 0.5 \right), \quad \omega, \phi \in [0,1]
\]  

(2)

To combine the two phases and benefit both from global uniqueness and high accuracy, \(\omega\) is first transformed to correspond to a position expressed as a multiple of \(\lambda\) by relating it to the projection width \(w\) and dividing by \(\lambda\). The non-integral remainder is then removed by subtracting \(\phi\), now that the two phases are expressed in the same unit, and the result is rounded to the next integral multiple of the wavelength \(\lambda\). Finally, the accurate phase \(\phi\) is added, and the projector coordinate \(x\) is obtained by multiplying with the lens-shifting wavelength. Subtraction of the phases in a common unit rather than more intuitively just rounding the coarse phase is necessary since especially at boundaries of periods, image noise can lead to erroneous classification of projector plane IDs. Now that for each camera point the absolute projector coordinate is known, 3D plane/ray intersection between projector planes identified by the projector coordinate and rays defined by picture positions results in 3D positions of the surface points according to the basic idea of structured light reconstruction.

5. Optimization

Reconstruction is subject to several influences, for instance camera noise or variable average luminance over captured images. Several optimizations were introduced to make reconstruction robust and more accurate.

Reconstructing a planar surface revealed instabilities of the numerical optimization used for fitting of the model function, resulting in the outliers visible in Fig. 5 (left). These problems were addressed by dividing the optimization process into two steps that differ by the set of parameters defined to be fixed while letting the others be subject to optimization. In the first stage, the fitting process is done by optimization with only the spatially varying phase being variable, while the frequency remains fixed and the amplitude is set to \(a = 1\) due to normalization. The resulting phase is then used as more accurate starting value for the second step with the remaining parameters, including the phase itself, being subject to optimization and thus variable, achieving a robust optimization as shown in Fig. 5 (right).

5.1. Image normalization and luminance profile smoothing

Even though the temporal change of luminance over a full lens-shifter sequence shows nearly sinusoidal behavior, there are still deviations, recognizable as depicted in Fig. 3 (upper left). The vertical highlight bar shows that deviations occur simultaneously for all pixels at the same lens-shifter position, captured by the same image. Since the phase is spatially dependent, this cannot be an effect of possible deviation from linearity of the lens-shifter, which would also contradict the linearity test conducted (see Section 3.2). Nor can it be a consequence of image noise that is spatially independent. The reason is found within a captured image compared to neighboring frames in time. The change of luminance over the course of frames deviating from the expected sinusoidal behavior is apparently caused by a per-frame bias in overall luminance. This is an effect of the consumer camera, which does not show precise repeatability, and of the missing synchronization with the oscillating projector signal. As a solution, the captured sequence is normalized to the average luminance of the first image in

---

Fig. 4. Luminance of pixel over a complete lens-shifting sequence and fitted model function.

Fig. 5. Reconstruction of a planar surface before (left) and after (right) 2-step fitting.
the sequence. This yields a significant improvement, as visible when comparing the upper right and lower left charts in Fig. 3. The final optimization that leads to a behavior sufficiently close to the target model function is filtering the luminance profiles of each pixel with a box filter of size 5 pixel (cf. the lower right chart in Fig. 3). After this step, the fitting process of captured data to the model function is robust, and reliable parameters are determined automatically.

6. Results

As our acquisition system targets the two domains of high accuracy digital preservation of cultural heritage objects and acquisition of height fields for material rendering, we show results of the two domains and provide an accuracy comparison between our approach and the pure phase shifting technique. An analysis of depth accuracy supports that the acquisition accuracy is indeed better than 100 μm, and tries to give an upper limit on depth accuracy beyond that limit. The same analysis is performed on a commercial scanning system for comparison.

6.1. Meso-scale reconstructions

In Fig. 8 (left) we show the rendering of a reconstructed leaf under green illumination. The leaf sets itself apart from the planar background geometry, jumping out into the third dimension. Note that we explicitly did not render true color, but used virtual illumination to highlight the fine geometry. The close-up shows a wireframe sub-region of the rendering, outlining the accurate reconstruction of the fine leaf branch structure. The reconstruction of a different leaf is pictured in gray-scale shaded rendering under white illumination in Fig. 8 (middle). The deep meso-structure ‘valleys’ formed by the fine leaf cell structure are plastically visible. White paint was used to exclude specular highlights during reconstruction.

The reconstruction of two objects from the cultural heritage domain, a 1955 Six Pence coin (silver) and a 1899 One Penny coin (bronze), are shown as renderings in Fig. 6. Despite the strong reflectance due to the shiny metallic material, no prior treatment was applied before reconstruction; yet fine details can be seen in the results (see Fig. 7 for a comparison of some common object preparation techniques). Small letters and complex details of the coined pattern are visible in the reconstruction of the 19 mm Six Pence coin. The One Penny comes at an estimated maximum coining elevation of 100 μm, with the major part of the coin having been ground down to zero over time. Only the right arm and torso of Victoria are left at maximum elevation, with a totally smooth transition into the zero-level coin background. These features and their smooth transition are captured well in the rendered reconstruction. Even though the year letters and the ‘ONE’ inscription are left at estimated 50 and 20 μm elevations, respectively, both coining details are still measured and reproduced by our scanning system.

6.2. Height field and BRDF material probe

Due to its meso-structure resolution, the acquired geometry can be used to apply a given BRDF material sample to the whole object surface. Fig. 8 (right) shows a ray-traced rendering of a 1 × 1 cm section of the leaf scan (middle) with a BRDF material sample of red car paint applied under illumination by an environment map. The rendering was generated by the ray tracer developed by Huff et al. [21] which is used to evaluate the BRDF model over the reconstruction surface. By evaluating the BRDF sample on all surface points of our high precision, meso-scale
height field, the effect created that the leaf appears like having been treated with the same exact red car paint the sample is taken from. Material-specific effects such as highlights and specularities captured by the BRDF sample are widely exploited due to its application to the full geometry, and compelling material-specific effects are reproduced.

6.3. Comparison to pure phase shifting

Since our acquisition system applies two independent techniques, a comparison of our approach to pure phase shifting is easily done by disabling the lens-shifting functionality and thus enabling a direct comparison without changing the hardware setup. To objectively compare the accuracy of the two approaches in relation to the number of images captured, we introduce the following metric:

$$\Delta = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \|P_1(x,y) - P_2(x,y)\|^2$$

(3)

For two reconstructions $P_1, P_2$ of the same object acquired immediately in sequence, the mean quadratic deviation is computed for each pair of 3D samples $(P_1(x,y), P_2(x,y))$ and accumulated over the reconstructed range, which for the evaluation was 500 x 500 pixels. Since our acquisition system is calibrated such that one unit in the reconstructed model corresponds to the length of 1 m in physical object space, we express the results of the metric in μm. The metric expresses the positional deviation between two subsequent scans which is an important upper bound on accuracy due to the heavy influence of noise, impacting reconstruction indeterministically. The pure phase shifting method was first evaluated using three wavelengths at 7, 11, and 13 pixels, and capturing three images per wavelength at different phase shifts each, i.e., a total of nine images. In the next run, seven images per wavelength were acquired, which correspond to the maximum number of different shifts possibly determined by the smallest wavelength, leading to 21 input images. To compare the approaches, our lens-shifting technique was evaluated once for 15, 25, 50, 100, and 200 captured images at a wavelength of two pixels, which means that the lens-shifter was driven to as many positions distributed evenly within the theoretic lens-shifting range of 2048 steps. The additional nine images required for phase shifting do not influence the accuracy measure as they serve for global phase determination only.

Table 1 shows the evaluation results.

For the tested object (the back side of a bathroom tile), already 25 lens-shifting images in combination with only nine phase shifting images are sufficient for our approach to improve the accuracy according to the measure in comparison to the pure phase shifting technique with 21 images used. Increasing the number of lens-shifting images further leads to slight improvements in the accuracy metric. Note that we reach an optimum at about 50 lens-shifting steps (see also Fig. 9). This is of course highly dependent on the scanned object and the hardware setup. The resolutions of the camera and the projector can especially influence the accuracy of the phase fitting. Therefore, to obtain optimal results we would recommend determining the specific optimal number of steps for the setup used beforehand. Depending on the complexity of the object, shadowing and specular reflection can influence the overall value of the accuracy, too.

The qualitative results of a simple planar surface are visually compared in Fig. 10 which is a rendered image of the reconstruction divided into two parts. The left part shows the result as achieved by our extended lens-shifting technique with a spatially coherent transition to the right part that is the result of the pure phase shifting reconstruction. In the pure phase shifting result, stripe pattern artifacts are apparent that follow the pixel
structure of the projected image, leading to deviation from the sine wave luminance flow. Our extended lens-shifting method does not show these artifacts and is significantly more error resilient. This comes at the cost of longer acquisition time needed in comparison to phase shifting alone, due to the additional phase shifting image acquisition and the higher computational cost for fitting. Reconstruction time for the full acquisition range of 120 × 80 mm is about 14.5 min on an Intel Core2 Quad 64 bit system at 2.5 GHz with 8 GB RAM under Microsoft Windows Vista 64 bit, including the acquisition time of about 5 min for the typically 50 images (nine for phase shifting, 40 for lens-shifting, one for texture). Parallelization of the algorithm on a GPU or several CPUs is straightforward, however, and can largely reduce the time needed, since the computation of each 3D position is independent from the others.

6.4. Assessing reconstruction quality: the confidence metric

In order to measure the confidence of our reconstructed 3D points we calculate the fitting error for each luminance profile. We therefore sum up all squared distances between the luminance profile $L_i$ of the pixels and the corresponding prediction of the fitted model function:

$$E = \sum_i (L_i(x) - a_i \cdot \sin(\phi_i + 2\pi f_i x))^2$$  \hspace{1cm} (4)

The quality is then measured as $1 - \sqrt{E \cdot t}$, where $t$ denotes a certain threshold (empirically set to 900 in our tests). For all errors above $1/t$ we clamp the quality to 0 to ensure that it lies in the interval $[0, 1]$. The results for this metric are shown in Fig. 11. We observe a low confidence especially at edges with shadowing artifacts and at regions of high specular reflectance. Here the luminance profiles are of course noisy and a good fit is often impossible. This is again strongly dependent on the scanned object and the hardware setup.

6.5. Breaking the 100 μm mark

Measuring depth accuracy is a complex task in such small dimensions. To quantitatively support the impression of high accuracy resolution given by the reconstructed coin profiles in Figs. 6 and 7, we constructed a ramp of paper sheets (5staroffice Re-Move notes). The measured height of a 100 sheet block allows the conclusion that the thickness of one sheet is approx. 93 μm. Shifting the sheets apart and placing the resulting ramp on the acquisition surface allows reconstruction of steps going into depth at a known height of 93 μm each. The rendering of the ramp reconstruction (Fig. 12, top) reveals that our scanning system is able to reconstruct the fine steps. Fig. 12 (bottom) is a height profile plot over a length of about 1 cm along the ramp. The steps are clearly distinguishable and at significant distance from the extent of noise over the sheet area, which is caused by the uneven micro-structure of the paper and impacted by camera noise.

This observation lets us assume that the depth accuracy is even beyond 100 μm. To determine an upper limit on depth accuracy, we constructed a continuous ramp going into depth. We used a steel bar with square cross-section and placed it on two pivots, resting on the scanning surface over the full length of the measuring range. The flat side of the steel facing the camera was covered with a white plastic layer from a camera color calibration chart. In this setting, we reconstructed the white surface between...
the two pivots. Then we elevated the steel bridge at one pivot by placing one of the sheets with known height between steel and pivot, and again reconstructed the white surface. When comparing the two reconstructions, there are two surfaces intersecting at the first pivot’s position and deviating in depth at the second pivot’s position (by 93 µm). In between, we thus achieve continuous depth differences between 0 and 93 µm. The position where the difference of the two height profiles sets itself apart from the noise leads to a minimum depth difference distinguishable by our 3D scanner, due to the known distance between pivots and height difference at the second pivot, and thus to an upper bound on its depth accuracy. Fig. 13 (top) visualizes the height difference between the profiles through the two reconstructions, for the region between the fixed pivot and the middle of the bar. A fitting line shows the linear profile trend and makes the magnitude of deviations clear. Already at 50% of the bar, deviations are small compared against the height difference of the two profiles. As this position corresponds to an object height difference of about 46 µm (half of the height difference achieved by elevation), it is obvious that depth accuracy is at least as high as 50 µm, which is a new lower bound on accuracy with still a high safety tolerance. Deviations from the expected surface geometry are largely caused by noise encountered in the optical sensor. This causes the apparently random and alternating deviation in either direction above or below the ideal profile. It is thus clear that the maximum resolution achievable in scanning depth is bounded by the maximum geometric deviation encountered over the scanning range, which in this test is at about 10 µm. Even with double safety tolerance, the position where the profile height difference exceeds twice the maximum deviation is reached at about 21% of the length, which leads to an estimated maximum depth accuracy of 20 µm. However, deviations shown here apply to reconstruction of the white material used and vary in magnitude for different materials, meaning that depth accuracy is dependent upon materials scanned.

6.6. Comparison to a commercial scanner

The very same test was conducted on a commercial industry high end scanning system (breuckmann smartSCAN³D-HE, M-300) that purely works on multi-period phase shifting and uses two CCD cameras at 5 MPix, with a baseline of 470 mm and a triangulation angle of 30°. The measuring range is 240 × 180 mm and the average distance to the object is 150 mm. With our system constructed from off-the-shelf-components, using only one camera at lower effective resolution, we achieve a comparable result. A close look at the comparison in Fig. 13, where the upper graph is the one already introduced for the test result on our system, while the lower graph presents the results for the commercial scanning system, even reveals that the geometric deviation shows less volatility and a lower maximum deviation than the system we compared against.

In addition to the quantitative comparison that focuses on depth accuracy, we show a qualitative comparison at the example of the One Penny coin, a complete rendering of which is shown in Fig. 6. We conducted a scan of the coin using the above commercial scanner which allows for qualitative high-detail comparison between the results of the two scanning systems for the same object and orientation. For this purpose, the respective reconstructed models were aligned using Iterative Closest Point (ICP) mesh registration prior to comparison. Fig. 14 reveals the differences in geometric details down to the triangulated structure of each of the scans. The software of the commercial scanner rejected scanning the untreated coin, which made the application of white spray paint necessary, while we were able to scan the object without any prior treatment using our scanning system. This explains the presence of a few holes in the reconstruction at positions with specular reflexes due to the reflective material. For the Six Pence coin, however, our result in contrary shows even fewer holes despite not using spray paint.

When comparing the two full reconstructions, the difference in detail stands out. While our result shows clear edges, especially in the region of the year coining, the result of the commercial scanner appears rather as having been smoothed. However, there is significantly less noise in the commercial solution than in our result where the surface is homogeneously covered by random noise. In the close-up of the arm, both effects become more clear. While our system reconstructs the weak, but present edge along the arm in a sharp way, the commercial solution appears to exhibit a soft and gradual incline that differs from the real object geometry. Noise can be seen again in our solution which now takes on the shape of a stripe pattern. This is likely caused by the projection of the high frequency stripe pattern which, due to the hardware implementation and its limitations, contains lines between the projected white and black columns that are of either red or green color and influence the fitting result. The wireframe-
view reveals that the commercial scanner software performs some postprocessing, like remeshing and smoothing. This causes sharp details like the center of the blossom or the “D” in the upper right corner of the wireframe closeup in Fig. 14 to get lost. Our scanning setup uses a regular grid (defined by the camera pixels) as surface mesh base. And again it is visible that the result contains sharper and stronger edges but also exhibits a higher amount of noise.

7. Conclusion and future work

We developed a meso-scale acquisition system targeting two domains, one being the high accuracy, fine scale and comparatively large range digital preservation of cultural heritage objects, the other one being the support of realistic rendering with measured meso-structure height fields of the materials to be reproduced. Using a 1 DOF mechanical lens-shifter extension allows reaching new levels of depth accuracy while avoiding limitations known from common structured light approaches or artifacts introduced by projection devices. A very low positional error is achieved, as the evaluation of deviations in the geometry resulting from two consequent scans showed, which is significantly below the one of the comparison technique. The area resolution of 55 μm is bounded by the camera used, while the depth resolution is influenced by camera noise and the projection technique. Our acquisition setup uses off-the-shelf-components and delivers high accuracy at low cost. Being bounded by the precision of hardware, the accuracy of our system could further be increased by gray scale cameras without SLR, avoiding vibration due to mirror movement and coming at higher luminance sensitivity and lower noise, let alone the higher capturing speed, or by adding a second camera which would increase precision and help in regions of high specularities not visible to the first camera. While we currently use a simple calibration model with manually measured parameters, not including lens distortion or similar effects and abstracting from lens-shifter motion, an automatic calibration procedure using a test pattern would be a valuable step further to avoid geometry distortions. Using a laser-based projector like the ultra-miniature, scanning projection Microvision PicoP projector would raise the accuracy due to sharp projection and avoiding dark regions between pixels. With miniature size and no need for focus adjustment, minimum distance to the object could be decreased, increasing the projected resolution and thus the area resolution of the reconstruction. Also, limitations on depth of field by projector and camera specifications can be ameliorated by miniature components or components with higher depth of field, thus increasing the current maximum acquisition depth of about 3 cm caused by the close positioning of the optical devices to the object.
Acknowledgments

This work was partially supported by the European projects 3D-COFORM (FP7-ICT-2007.4.3-231809) and MAXIMUS (FP7-ICT-2007-1-217039) and the DFG Emmy Noether fellowship GO 1752/3-1.

References


Fig. 14. Comparison between scans of One Penny and Six Pence coins conducted with a commercial scanner (left) and with our scanning system (right). 1st and 2nd row: full gray-scale rendered view and subsection with arm and ‘one’-inscription for One Penny. 3rd and 4th row: full gray-scale rendered view and subsection with blossom for Six Pence. 5th row: triangulation close-up of Six Pence blossom.


