Scene-adapted Structured Light

Thomas P. Koninckx[1]   Pieter Peers[2]   Philip Dutré[2]   Luc Van Gool[1,3]

[1][2] Katholieke Universiteit Leuven,   [3]Swiss Federal Institute of Technology (ETH)
[1] Dep. ESAT / VISICS,[2] Dep. Computer Science   D-ITET / BIWI,
Leuven, Belgium   Zürich, Switzerland
{tkoninck, vangool}@esat.kuleuven.ac.be ,   {pieterp, phil}@cs.kuleuven.ac.be

Abstract

In order to overcome several limitations of structured light 3D acquisition methods, the colors, intensities, and shapes of the projected patterns are adapted to the scene. Based on a crude estimate of the scene geometry and reflectance characteristics, the local intensity ranges in the projected patterns are adapted, in order to avoid over- and under-exposure in the image. This avoids the infamous specularity problems and generally increases accuracy. The estimated geometry also helps to limit the effect of aliasing caused by the sampling of foreshortened patterns. Furthermore, the approach also acounts for the adverse effects that small motions during scanning would normally have. Moreover, the approach yields a confidence measure at every pixel of the range image. Last but not least, the scanner consists of consumer products only, and therefore is cheap.

1. Introduction

1.1. Rationale

Noncontact optical surface digitization techniques have evolved substantially during the last decades [1]. Structured light techniques, based upon the projection of 2D patterns [2], form an interesting subset. One of the more advanced instances combines a time series of binary or color coded patterns with a limited number of shifted 2D sine waves [3]. The former yield robust but low resolution 3D information. The latter, through a technique often referred to as *profilometry* or *phase shifting* [4], are used to boost the local resolution. We follow a similar strategy, but with patterns that self-adapt to the scene.

Structured light techniques offer interesting features. The cost is low, as consumer hardware is used (a combination of cameras and projectors), and there are no fragile, moving parts. This also yields higher speeds. Eye-safety is no issue, and the setup is usually easy to reconfigure. This said, there are also some serious limitations. Specular objects or scenes with very bright and dark parts may lead to dynamic ranges that the camera(s) cannot handle. As a consequence, parts of the patterns will go undetected or are misread. Also, the projected fringe patterns may appear foreshortened to the point where the camera can no longer correctly pick up their high, spatial frequencies (aliasing). Also, the scene and system are assumed to be perfectly static while the series of patterns are projected. Blais et al. [5] noted that this may be unexpectedly hard to achieve in practice. Finally, nonlinearities in both cameras and projectors are most often completely ignored, again leading to distorted models. Even worse, all but the saturation problem risk to remain undetected. These problems are often addressed through more conservative system design, at the cost of rendering the devices slower, more expensive, and bulkier. Even then, quite some manual intervention may be required, e.g. spraying or powdering the objects.

What we envision is a system which rather adapts itself to the scene. Ranges are acquired in a two step procedure. First the scene is analyzed. This comprises a crude estimation of scene geometry and measures to take surface reflectance into account. Then, the projection patterns are adapted to avoid under- or over-exposure, as well as aliasing problems. This leads to an iterative process, where the deviation between expected and observed patterns allows the system to refine the geometry and provides it with confidence measures. The process stops when neither the geometry nor the confidences are updated. The estimated geometry also facilitates the detection and compensation of motion between consecutive frames. On the photometric side, nonlinearities and crosstalk between the color channels are modeled for both the camera and projector. The outcome still is a very low cost system, based on one normal camera and one LCD projector, that now can deal with a wider category of objects (e.g. metal industrial workpieces) and that requires fewer interactive settings.

1.2. Related work

The work of Caspi et al. [6] is one of the first structured light approaches which explicitly models the camera-projector path. The main purpose however is robust color coding. More recent work on color coded structured light [7] focuses mainly on limiting the number of projection planes

for fast 3D. A four shot real-time structured light approach was suggested in [8], and in [9] single-shot 3D with adapted patterns is used. In the seminal work of [10] the question of an optimal set of patterns is formulated for the first time. Online feedback to the scene however is not yet included, and conclusions about good patterns are applied uniformly to the entire scene, not locally. We propose a more complete projector-camera model which will allow to adapt patterns online and on a per pixel basis. This is a bit similar to the work proposed in [11] in which a projector is used to change an object's appearance. However, only very crude geometry is used, and no difference between reflectance and color crosstalk seems to be made.

Parts of the work here is akin to recent advances in high dynamic range imaging. Recently Nayar et al. [12] implemented such process by looking *through* a modified data projector. Albeit for completely different applications, this demonstrates the power of a per pixel intensity modulation. Various techniques for high dynamic range imaging and the estimation of response curves have been proposed. See e.g. [13, 14] for an implementation working on stills, or [15] for streaming video.

The rest of the paper is organized as follows: section 2 deploys the camera-projector model and the resulting technique to overcome camera saturation, while section 3 discusses the pattern geometry. In section 4 we explain how patterns are decoded and how motion during the acquisition is taken into account. Results are shown in section 5. Section 6 concludes the paper.

2. Active Dynamic Range Compensation

This section models the radiometric chain from projector values to image intensities. This model will be used to boost the precision of profilometry, by avoiding saturation in the camera image.

2.1. Trichromatic projector-camera model

Suppose we illuminate a scene with M projection patterns. The nth 'pixel' of the mth, $m = 1 \ldots M$ pattern will be be denoted as $x^p_{n,m,b}$, with the superscript p referring to the projector and $b \in \{R, G, B\}$ to the color band. $\mathbf{x^P_{n,m}}$ is the corresponding R,G,B-triplet. A similar notation $\mathbf{x^c_{n',m}}$ is used for the n'th pixel in the camera. In this section, the relation is derived between the applied projector color signal $\mathbf{x^P_{n,m}}$ and corresponding camera image color values $\mathbf{x^c_{n',m}}$. An overview of the different steps is shown in fig. 1.

Cameras typically respond nonlinearly to the incoming irradiance. Similarly, the radiant flux generated by the projector depends nonlinearly on the signal that is applied to its different 'pixels'. These nonlinearities tend to boost perceptual quality, but complicate profilometry which needs to
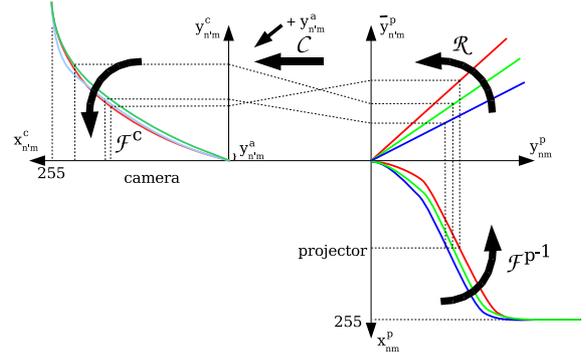

Figure 1: *The projector-camera model. Bottom right: the projection pattern is transformed by the inverse projector response curve. Top right: reflection by the scene. Top middle: crosstalk between the color channels of the camera and the projector Top left: from irradiance to pixel values via the camera response function.*

take them into account. In the sequel we will denote the effective radiant fluxes for the three color bands at pixels n and n' as $\mathbf{y^P_{n,m}}$ and $\mathbf{y^c_{n',m}}$, resp. The nonlinear response curves for both camera and projector map these fluxes to a discrete set of pixel values, and this for the three color bands. For an 8 bit image:

$$\mathcal{F}_b : \mathcal{R}^+ \to [0, 255], \quad b \in \{R, G, B\}. \tag{1}$$

The first step in our model therefore is the transition of projector pixel values $\mathbf{x^P_{n,m}}$ to the resulting radiant fluxes $\mathbf{y^P_{n,m}} = \mathcal{F}^{\mathbf{P}-\mathbf{1}}(\mathbf{x^P_{n,m}})$.

Of course, the scene will modulate how much of the radiant flux coming from the projector will actually reach the camera lens. Of this flux $\mathbf{y^P_{n',m}}$ only part will be reflected towards the camera. We assume a simple, linear reflectance model. At each 3D point $[X, Y, Z]$ of the scene these reflected fluxes are obtained as:

$$\mathbf{\bar{y}^P_{n',m}} = \mathcal{R}(\mathbf{y^P_{n,m}}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) =$$
$$\begin{bmatrix} r_{1,1(X,Y,Z)} & 0 & 0 \\ 0 & r_{2,2(X,Y,Z)} & 0 \\ 0 & 0 & r_{3,3(X,Y,Z)} \end{bmatrix} \begin{bmatrix} y^p_{n,m,R} \\ y^p_{n,m,G} \\ y^p_{n,m,B} \end{bmatrix} \tag{2}$$

Although both the spectral channels of projector and camera have been referred to as 'R', 'G', and 'B', their spectral compositions will by no means be identical. This 'mismatch' between generated and recorded light is modeled by a crosstalk function $\mathcal{C}$:

$$\mathbf{y^c_{n',m}} = \mathcal{C}(\mathbf{\bar{y}^P_{n',m}}) + \mathbf{y^a_{n',m}}. \tag{3}$$

For $\mathcal{C}$ a quadratic model is chosen in contrast to the usual first order model. This choice has been experimentally justified, as will be shown later. As we don't want to work under

2

darkened conditions, also a per pixel ambient light contribution $\mathbf{y^a_{n',m}}$ will reach the camera, which is also added at this stage of the model.

The last step translates the incoming radiant flux $\mathbf{y^c_{n',m}}$ at the camera back into the pixel intensities $\mathbf{x^c_{n',m}}$ for the three camera color bands, via the camera response function: $\mathbf{x^c_{n',m}} = \mathcal{F}^c(\mathbf{ty^c_{n',m}})$, with t the exposure time of the camera.

2.2. Calibrating the projector-camera chain

In what follows we need explicit knowledge of $\mathcal{F}^c$, $\mathcal{F}^p$ and $\mathcal{C}$. This calibration step should be done only once, and can be computed offline. $\mathcal{R}$ and the geometry are scene dependent and therefore part of the online process.

Determination of $\mathcal{F}^c$, $\mathcal{F}^p$: In our experiments we will use the technique proposed by Debevec *et al.* [13] to determine $\mathcal{F}^c$. It requires a set of images of a static scene with a varying but known integration time.

For the determination of $\mathcal{F}^p$, a similar approach can be followed, as a projector is an 'inverse camera'. A complicating factor is that the relation between the projected patterns (i.e. between the radiant fluxes) needs to be analyzed via the camera. Hence, we have to work our way back from camera observations $x^c_{n',m,b}$ to $y^p_{n,m,b}$. This requires going via $\mathcal{F}^{c-1}$, C^{-1}, and R^{-1}. $\mathcal{F}^c$ is known already and R we can basically eliminate by projecting uniform patterns onto a white, planar, uniform, and diffuse surface. We have used 30 patterns, each uniform but of 10 different intensity levels $i^p_{m,b}$ and activating only one of the three color channels of the projector at a time. Each pattern was only observed in the corresponding camera color band. By performing the analysis for each color channel separately, we have tried to minimize the effect of not taking C into account as it is not known yet. For each of the patterns, we have avoided to saturate the camera by adjusting the integration times t_m accordingly. Undoing the camera nonlinearity by $y^c_{n',m,b} = \mathcal{F}^{c-1}(x^c_{n',m,b})$ and correcting $i^p_{m,b}$ for the differences in integration times t_m yields an approximation of the radiances. From there, discarding the influence of $\mathcal{C}$, we apply the same procedures for the projector as used for the camera. In the next paragraph $\mathcal{C}$ will be determined, and we could in fact iterate through repeated $\mathcal{F}^p$ and $\mathcal{C}$ estimations. As corroborated later, this was not necessary.

Fig. 2 shows the result for our Sharp XGC55x data projector, and AVT Marlin F080C IEEE1394 camera.

Determination of $\mathcal{C}$: In order to estimate $\mathcal{C}$ we add to the earlier set of images a series of shots with illumination in all three of the color channels so that a total of $10 \times 10 \times 10$ combinations of intensity levels are applied. This yields a uniform sampling of the color space. The projection patterns and the corresponding radiance maps of the images minus the ambient light $\{\mathbf{y^p_{n,m}}, \mathbf{y^c_{n',m}} - \mathbf{y^a_{n',m}}\}$ are used to estimate a least squares solution which maps both datasets
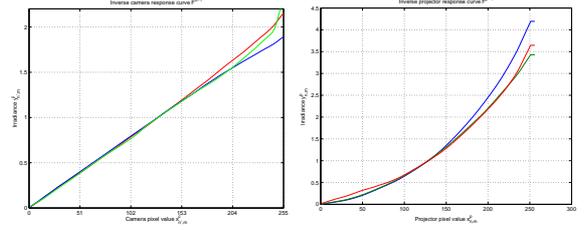

Figure 2: *Camera and projector response curves. Left: camera; close to linear. Right: projector; clearly nonlinear behavior.*

to each other. Eq. (4) gives the result as $\mathcal{C}^{-1}$ (camera $\rightarrow$ projector):

$$
\begin{bmatrix} \bar{y}^p_R \\ \bar{y}^p_G \\ \bar{y}^p_B \end{bmatrix} = \begin{bmatrix} 1.52 & 0.11 & 0.08 \\ 0.20 & 1.24 & 0.16 \\ -0.06 & 0.09 & 1.43 \end{bmatrix} \times \begin{bmatrix} y^c_R \\ y^c_G \\ y^c_B \end{bmatrix} +
$$
$$
\begin{bmatrix} 0.15 & -0.01 & 0.13 \\ 0.14 & 0.10 & 0.48 \\ 0.03 & 0.03 & 0.37 \end{bmatrix} \times \begin{bmatrix} y^c_R y^c_G \\ y^c_G y^c_B \\ y^c_B y^c_G \end{bmatrix} +
$$
$$
\begin{bmatrix} -0.24 & -0.36 & -0.16 \\ -0.23 & -0.44 & -0.45 \\ -0.05 & -0.42 & -0.32 \end{bmatrix} \times \begin{bmatrix} y^{c\,2}_R \\ y^{c\,2}_G \\ y^{c\,2}_B \end{bmatrix} \quad (4)
$$

The relative size of the second row of the first order coefficients indicates that crosstalk from red and blue to green is considerable. Also the second order coefficients cannot be neglected. We have tested for higher order models, but the corresponding components were negligible.

Dealing with $\mathcal{R}$: In order to perform profilometry safely, one has to avoid getting into the saturation regions – dark and bright – of the camera. It is therefore important to determine which input levels at the projector side give rise to such conditions, i.e. we need to know the intensities at every pixel in the projector $\mathbf{y^{Psat+}_n}$ and $\mathbf{y^{Psat-}_n}$ which will generate these saturation levels, for resp. over and under exposure. As this analysis is dependent on the scene geometry and reflectance properties (summarized as $\mathcal{R}$), it is part of the online, 3D acquisition process.

Hence, before the projection of the patterns used for the 3D acquisition, several patterns are projected first, specifically with this goal in mind. These are uniform, gray patterns of different intensity. We used three, with an intensity of resp. 10%, 50% and 90% of the projector's maximal output. This yields a triplet $\{\mathbf{y^c_{n',1}}, \mathbf{y^c_{n',2}}, \mathbf{y^c_{n',3}}\}$ that reach the camera. For the sequel, it is important that at least one of these three values keeps the camera out of saturation. For over exposure, these levels can be calculated for the three spectral channels as $\mathbf{y^{csat+}} = \mathcal{F}^{c-1} \begin{bmatrix} 240 \\ 240 \\ 240 \end{bmatrix}$. Note that we took a small safety margin with respect to the true maximal level of 255. This can be translated into maximal levels $\bar{\mathbf{y}}^{Psat+}$ in each of the projector's spectral channels by lifting the crosstalk: $\bar{\mathbf{y}}^{Psat+} = \mathcal{C}^{-1}(\mathbf{y^{csat+}})$. Similarly, starting from a safe minimal level of 15, the lower thresholds $\bar{\mathbf{y}}^{Psat-}$ are determined.

3

In the previous paragraphs we have modeled the nonlinearities in the projector-camera radiant flux transfer. This knowledge will be exploited now. Given an observed relation between $\mathbf{y}^c_{n',m}$ and $\mathbf{y}^p_{n,m}$ – obtained directly from the camera readings and projection pattern values through $\mathcal{F}^{c-1}$ and $\mathcal{F}^{p-1}$ resp. – we can compute the linearized relation between $\bar{\mathbf{y}}^{\mathbf{p}}_{\mathbf{n'},\mathbf{m}} = \mathcal{C}^{-1}(\mathbf{y}^{\mathbf{c}}_{\mathbf{n'},\mathbf{m}})$ and $\mathbf{y}^p_{n,m}$. The ratio between these values -being the reflectance $\mathcal{R}$- should be constant for a fixed image pixel over all patterns m and fluxes $\mathbf{y}^p_{n,m}$ as long as it is correctly exposed. It is now an easy matter to compute the saturation levels $\mathbf{y}^{psat+/-}$ from $\mathcal{C}^{-1}(\mathcal{F}^{c-1}(240))$ and $\mathcal{C}^{-1}(\mathcal{F}^{c-1}(10))$. For the sake of precision, this process is repeated for multiple patterns and a weighted least squares regression yields the final critical projection levels $\mathbf{y}^{psat+/-}$.

Knowledge about the nonlinearities and saturation levels will be used in the next section for per pixel optimisation of the mean and amplitude of the sine patterns used in profilometry. Through the use of the multiple projection patterns and our radiometric camera calibration, we also get a high dynamic range texture map at no extra cost.

So far, we have glossed over the correspondence between projector pixels n and camera pixels n', in order not to unnecessarily complicate the discussion. Indeed, this is not an issue as long as the projection patterns are uniform, as was the case in this section.

3. Geometry Driven Patterns

In the previous section, we have derived what the acceptable range of pattern values is at the projector in order to avoid saturation in the image. This knowledge will be exploited for profilometry here. This section also deals with a similar, geometrical issue: how to choose acceptable projector pattern frequencies to avoid aliasing in the camera image. This will be a fully online process, as again scene geometry will play a pivotal role in this pattern deformation from projector to camera. These considerations lead to a reversal of what is usual: starting from what would be ideal patterns for the camera to accurately pick up – uniform sine waves centered around 128 with maximal amplitude and with a period between 10 to 20 pixels – the projector patterns are to be designed so as to yield these very image patterns.

Camera-projector pattern transfer: In optical profilometry a series of L phase shifted 2D sine patterns is projected. The m'th pattern is imaged by the camera as:

$$x^c_{n',m} = A_{n'} + B_{n'} cos(\phi_{n'} + m\delta), \qquad (5)$$

with $A_{n'}$ the albedo, $B_{n'}$ the sine amplitude of pixel n', $m = 1 \ldots L$, and $\delta = \frac{2\pi}{L}$ the phase shift.

Minimization for every pixel n' of:

$$\sum_{m=0}^{L} (x_{n',m} - A_{n'} - A'_{n'} cos(m\delta) + A''_{n'} sin(m\delta))^2, \quad (6)$$

yields the unknowns $A_{n'}$, $A'_{n'} = B_{n'} cos(\phi_{n'})$ and $A''_{n'} = B_{n'} sin(\phi_{n'})$. The phase $\phi_{n'} = atan\left(\frac{A''_{n'}}{A'_{n'}}\right)$ corresponds to a subpixel location in the pattern, and as such solves the correspondence problem for every pixel up to an unknown number of sine periods. These offsets can be computed by phase unwrapping for continuous surfaces [4], by the use of multiple periods [3] or by a binary/Gray coded sequence preceding the sine waves [2].

It remains however unclear which period to use in the design of the sine patterns. Fig. 3 illustrates the problem. A pattern projected from point B onto the right plane is foreshortened when observed from a viewpoint A. There is a risk of observing distorted patterns both due to aliasing because of the discrete sampling in the camera (picking up wrong frequencies) and because of integration over finite pixels (frequencies being washed out). In the literature this problem has been dealt with by hand-picking a conservative choice for the highest projected frequency. Here, we propose pattern adaptation. A location specific threshold for distortions to start can be computed given approximate knowledge of the geometry. To obtain the latter, a Gray coded pattern sequence is used after the three online test patterns of the previous section. As a matter of fact, the dark and bright one of these would normally be part of Gray coding and don't have to be repeated. The Gray sequence is limited to low frequencies as only approximate geometry is needed in this bootstrapping stage.

With approximate geometry at our disposal, we can warp the camera image to the one which would have been 'seen' at the projector. But first the geometry is filtered to remove noise through a combination of median, bilateral, and morphological filters, and to fill holes, though thin-plate spline interpolation. Finally, the geometry is triangulated, such that warping can be carried out extremely efficient with the GPU. The camera to projector warping function will be denoted as $\mathcal{W}$. At this point, it is useful to note that we also take the radial distortions of camera and projector into account.
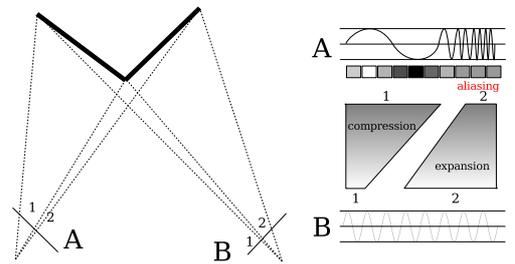

Figure 3: *Left: top view of a projector and a camera looking at a perpendicular jig. Right: aliasing of a sine pattern due to the foreshortening by the scene.*

Pattern generation: Given the warping function $\mathcal{W}$, a

pattern with desired geometry at the side of the camera can be generated and projected. As we want to apply profilometry, ideally the camera irradiances resulting from the camera should be sine waves around $\mathcal{F}^{c-1}(128)$, swinging between the saturation levels $\mathbf{y^{Psat-}}$ and $\mathbf{y^{Psat+}}$. In particular, the pattern values at the projector will take the form :

$$\mathcal{F}^p \left(\frac{\mathbf{y_n^{Psat+}} - \mathbf{y_n^{Psat-}}}{2} cos\left(\phi_n + m\delta\right) + \frac{\mathbf{y_n^{Psat+}} + \mathbf{y_n^{Psat-}}}{2} \right) \tag{7}$$

but before applying it, this pattern is warped by $\mathcal{W}$. As a result, at the side of the projector a pattern with *spatially dependent frequency and intensity* appears.

Referring again to fig. 3, the setup clearly is symmetrical. As much as we did improve the pattern for the camera, aliasing problems may now appear at the projector's side. Frequencies may get to high there to be generated. Therefore, the system works with a dual process. Apart from the deformed patterns, uniform sines are also projected, but still with the necessary intensity modulation of eq. 7. Aliasing for camera and projector will occur at different locations in the scene, thereby providing complementary information with the camera-based and projector-based sine projections. Also, as we can push more closely towards the resolution limits of both devices, the spatial resolution of our 3D reconstruction will increase.

The result of both series of sine waves are combined in a principled way. $\mathcal{W}$ and $\mathcal{W}^{-1}$ allow to predict the closeness to the Nyquist frequency at every location for each of the two schemes. This yields weights used in a linear combination of both reconstructions.

4. Decoding and Motion Compensation

After removing the influence of the response curve of the camera $\mathcal{F}^c$, equation 6 is used to compute the phase $\phi_{n'}$ at every camera pixel. These phase values will be used to update the initial geometry which was used to bootstrap the process. As we designed the projection pattern the way it should be reflected to the camera, the corresponding 'desired' phase $\phi_{n'}^*$ expected at each pixel is known. The sum of absolute differences between 'reflected' and 'desired' phase evaluated over a 7×7 window:

$$\Delta\phi_{n'} = \sum |\phi_{n'} - \phi_{n'}^*| \quad \forall n' \tag{8}$$

gives a measurement of the error on the initial geometry. See fig. 4, bottom row.

The translation for every pixel which transforms the decoded phase map into the expected phase map, is sufficient to update the geometry. This can be used to refine the correspondences between camera and projector as the gradient from geometry to phase is known analytically. We will

search this specific translation which minimizes simultaneously the difference between decoded and expected phase $\Delta\phi_{n'}$ (we assume the measurement to be close to correct) and the difference between the phase of a pixel and its neighbors (continuity). Referring to fig. 4, column 3 row 1 & 2: every 'band of equal phase' will become vertical but we stay as close as possible to the original solution. A graph cut algorithm [16] is used to solve this minimization problem. This results in an update of the geometry both for the parts with a good initial estimation as for the other parts. It is optional to use this updated geometry to generate a new warp function $\mathcal{W}$, regenerate the patterns and re-iterate the range acquisition. This iteration can be continued until geometry and error measurement remain unchanged. Alternatively one can threshold based on $\Delta\phi_{n'}$, and only retain the trustworthy parts of the reconstruction. Again referring to fig. 4 two steps of such an iteration are depicted. Geometry update and error reduction are clearly visible.

The decoding described above, asks for perfect immobilization of the object in between the consecutive frames. If this is not the case a sine wave modulation (see fig. 10) will appear on the resulting geometry. (A similar distortion occurs in case of over exposure, see fig. 5.) If we approximate the geometry locally as a planar surface, we know that the phase shifted patterns should be symmetrically interleaved. E.g. in case of 4 patterns every period of the 3 th pattern should start exactly in the middle of a period of the first pattern. In reality this will never be the case, because the assumption of planar underlying geometry is invalid. The deviation should in general cases however be random and non-systematic. This assumption can be used for detection and correction of motion in between the shots. For those parts of the phase map with low $\Delta\phi_{n'}$ the reflected sine patterns will be very close to vertical. The phase map in these areas is used to initialize vertical lines which correspond to the start of a period in each reflected sine image, and for every period. Those lines are trimmed to their subpixel location based on a matched filter. If the trimmed lines in the different sine images all show a similar translation, the object did move. A warp of the input images which applies the inverse translation and a re-estimation of all phase values ϕ greatly reduces the effect of the motion. It is self-explanatory that the camera-based patterns, which will show as vertical sines greatly facilitate this operation when compared to normal skewed patterns.

5. Results

Fig. 4 shows results and intermediate steps for two consecutive iterations of our algorithm. Note that during the first iteration the geometry between the label on the book and the rest of the scene is completely interpolated. This interpolation is rather accurate in the region under the label, which is not the case in the region next to it. During the second it-
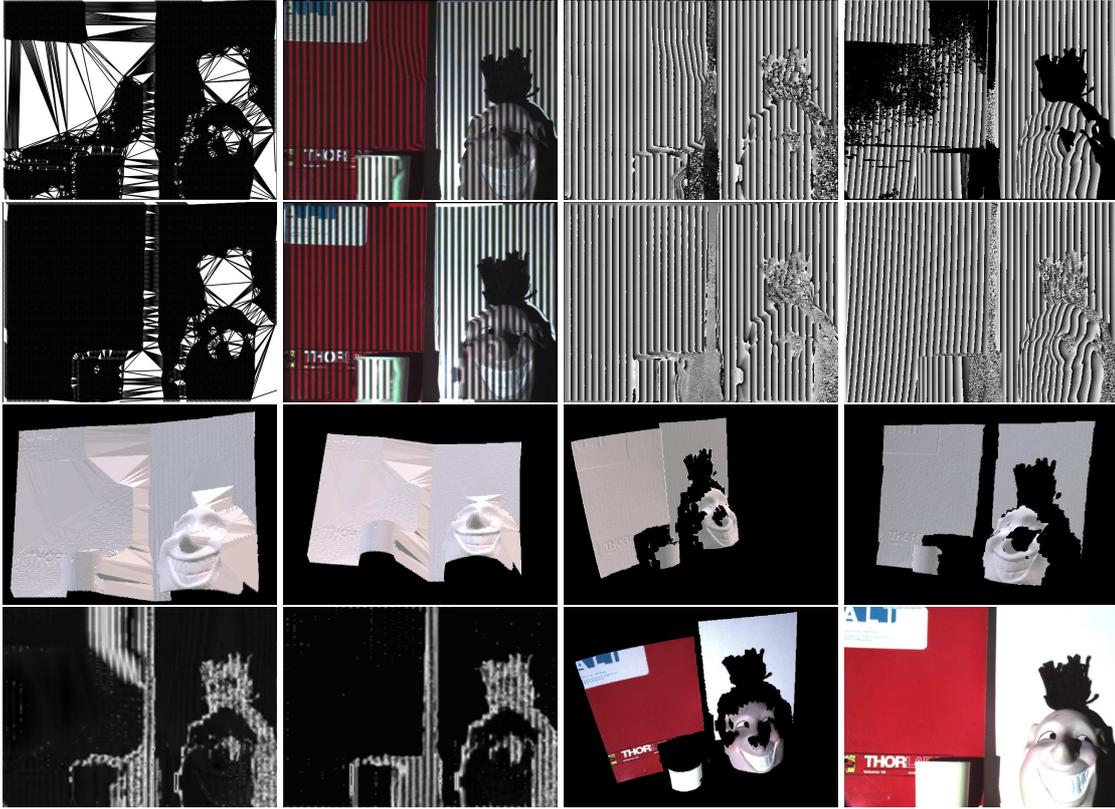
Figure 4: *Overview of the algorithm, during two consecutive iterations. Top row, from left to right: visualization of the initial warp from camera to projector, reflected camera-based sine pattern, decoded camera-based phase and the projector-based phase* without *use of the photometric model. Second row: same results as above after the second iteration. Third row: left: interpolated geometry used as input, right refined resulting geometry. Bottom row: uncertainty on the geometry after the first and second iteration, textured geometry, input image.*

eration real measurements replace the interpolated data (see e.g. the density of the warp in the first shot of the second row). The per pixel uncertainty, and the 'straightness' of decoded phase and patterns computed in the camera after both steps confirm this conclusion. The regions with high uncertainty after the second iteration are mainly caused by cast shadows. This can be verified in the image as seen from the camera in the bottom right. This last shot also illustrates the intensity range reflected.

Fig. 5 illustrates the effect of over exposure in the camera. The geometry presented to the scanner was a slightly specular planar surface. The circular artefact and holes in the top-left reconstruction are caused by the mirror reflection of the projector. All of the reconstruction suffers from a strong sine wave modulation caused by clipping of the reflected pattern. The top right reconstruction shows what our algorithm is capable of.

The model for the crosstalk function $\mathcal{C}$ was validated by a comparison between predicted and measured camera irradiances for a high number (4×255) of projected intensities and colors. The result is shown in fig. 6 and demonstrates a
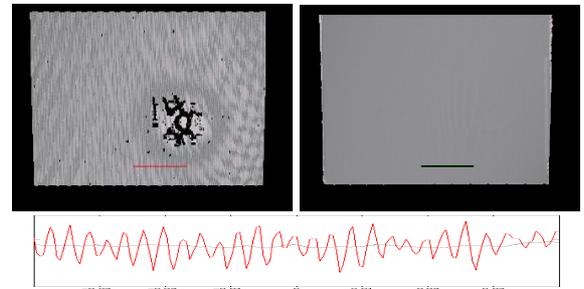

Figure 5: *Left: incomplete reconstruction due to camera over exposure. Right: the corrected planar geometry using our technique. Bottom: crossection for the line indicated (scale in this plot top to bottom: 1,5 mm).*

close fit between model and data. This test confirms that a linear model indeed would be insufficient (see also the discussion of eq. 4), and that the suggested iteration over the estimation of $\mathcal{C}$ and $\mathcal{F}^p$ is not necessary.

In fig. 7 the results of a test with a Macbeth colorchecker are shown. Our model was used to compute the projection
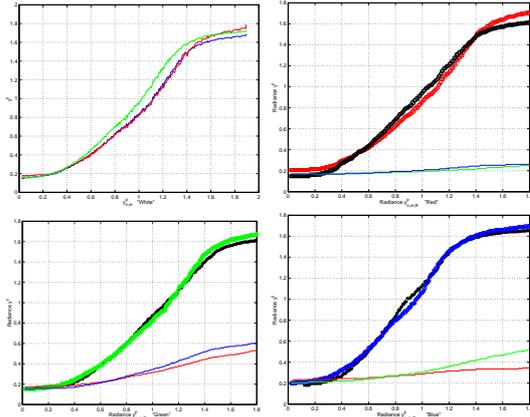
Figure 6: *Crosstalk measured vs. modeled. Top left: camera response for uniform illumination. Rest: camera response for resp. red, green and blue illumination by the projector. The thick black line shows the model fitted.*

pattern needed to get a uniform gray (20%) camera image. The histogram for every color channel shows a satisfying result, given the intense and saturated colors in the scene. Fig. 8 shows a similar experiment: the over- and under sat-
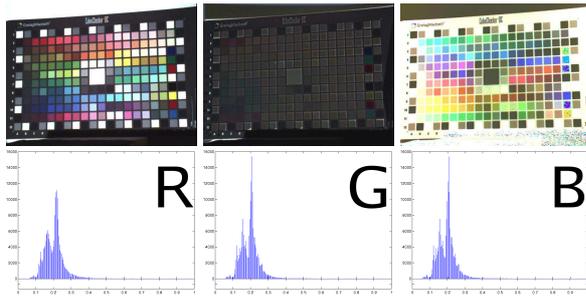

Figure 7: *Top: input image, 'compensated image' and 'compensation projection pattern'. Bottom: histogram for the red, green and blue channel.*

uration threshold levels are computed, and compared to the values measured when projecting all possible intensity levels. In fig. 9 a situation similar to the one of the illustration in fig. 3 shows how taking aliasing into account leads to a more complete reconstruction. Fig. 10 shows a preliminary result of our motion compensation strategy. The artefact on the reconstruction of the water can is greatly reduced. The technique proposed so far is used to make a complete reconstruction of several objects. Fig. 11 shows a set of 12 aligned shots of a 10 cm long goose statue. The two photographs on the right of fig. 11 allow to get an idea of the surface texture and the degree of specularity. For this object mainly better accuracy and a higher degree of fine detail (see e.g. the wings) resulted.

Fig. 13 shows a series of scans of a wheel made out of a specular alloy. Note that in both models the scans are only

Figure 8: *Top: projection intensities causing over and under exposure predicted by the model. Bottom: ground truth.*
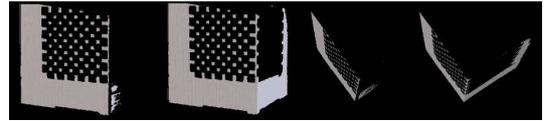

Figure 9: *Reconstruction of a calibration jig with and without taking aliasing into account.*

aligned and not integrated in a high quality model yet. Both models are generated using a single iteration for each scan. This also demonstrates that both the iterated and single pass application of the technique proposed is useful. In fig. 12 it is illustrated what happens if we try to scan the wheel without taking surface reflectance into account. The strong curvature makes that there will always be a mirror reflection which is cast directly into the camera.

The number of patterns which are added when compared to similar techniques remains very limited. The three test patterns needed for taking $\mathcal{R}$ into account only introduced one additional pattern, as the two others could be reused during the Gray coded sequence. The projection patterns added to minimize aliasing in the camera effectively doubled the length of the sequence needed for profilometry, which typically uses 4 shots. This brings the overall length of a complete sequence on 16 frames, which allow for acquisition at approx. 2 fps. The compactness of the sequence becomes evident when e.g. compared to the ca. 100 frames reported in [3], which are needed to take device resolution and scene reflectance into account by exhaustively testing all possibilities. In case no additional patterns can be tol-
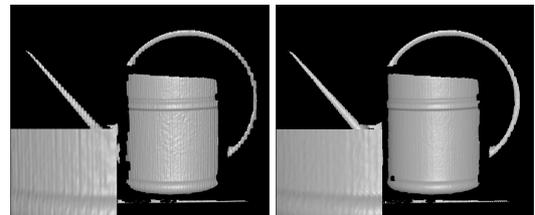

Figure 10: *Right vs. left: with vs. without motion compensation. The bottom left part of each figure shows a zoom-in.*
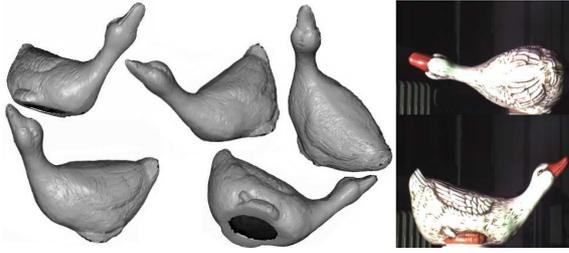
Figure 11: *A complete reconstruction of a small statue.*

Figure 12: *Left: reconstruction with and without taking surface reflectance into account. Right: uniform sine pattern and plane white illumination as seen from the camera.*

erated it can be still useful to work with the camera based patterns when compared to the more classical projector oriented approach, because the resolution of the projector is typically higher than the one of the video camera. This results in better handling of the aliasing problem.

6. Conclusions and Future Work

We demonstrated that fast and high quality modeling with a low cost setup is possible. By use of an extended camera-projector model nonlinearities in both devices, crosstalk and possible over- and under exposure in the camera are catered for. This is done online with only limited computational overhead, and without an excessive increase of the number of projection patterns. By generating camera based and projector based patterns we effectively minimize all aliasing effects in both devices. An uncertainty for every point of the reconstruction is computed, and distortions due to

Figure 13: *A reconstruction of a specular metal wheel.*

motion in between the phase shifted shots can be detected and partially corrected. The approach can be used both in an iterated and non-iterated way. This turns an inexpensive setup into a performant tool for range acquisition, and allows to reconstruct even fairly specular objects.

In this work surface reflectance properties are taken into account to minimize the influence on the acquisition of the geometry, and the texture are acquired in high dynamic range with a corrected color balance for all intensities. Ongoing research covers the explicit acquisition of the surface reflectance properties themselves together with the geometry.

Acknowledgments

References

[1] F.Blais, Review of 20 years of range sensor development, Electronic Imaging, 13-1, pp. 231-240, 04.

[2] J.Batlle, E.Mouaddib and J.Salvi, Recent Pogress in Coded Structured Light as a Technique to Solve the Correspondence Problem, Pattern Recognition, vol 31, nr.7, pp.963-982, 98.

[3] D. Scharstein, R. Szeliski, High-Accuracy Stereo Depth Maps Using Structured Light, CVPR, pp. 195-202, 03.

[4] M.Herraez, A.Gdeisat, D.Burton, M.Lalor, Robust, fast, and effective two-dimensional automatic phase unwrapping algorithm based on image decomposition, Applied Optics, vol. 41, nr. 35, 02.

[5] F.Blais, M.Picard, G.Godin Recursive Model Optimization Using ICP and Free Moving 3D Data Acquisition, 3DIM, pp. 251-259, 03.

[6] D.Caspi, N.Kyriati, J.Shamir, Range Imaging With Adaptive Color Structured Light, PAMI vol. 20, nr. 5, pp. 470-480, 98.

[7] C. Je, S. Lee, R. Park, High-contrast color stripe pattern for rapid structured-light range imaging, ECCV, pp. 95-107, 04.

[8] O. Hall-Holt, S. Rusinkiewicz, Stripe Boundary Codes for Real-Time Struct.-Light, ICCV, pp. 359-366, 01.

[9] T. Koninckx, A. Griesser, L. Van Gool, Real-time Range Scanning of Deformable Surfaces by Adaptively Coded Structured Light, 3DIM, pp. 293-302, 03.

[10] E. Horn, N. Kiryati, Towards Optimal Structured Light Patterns, Image and Vision Computing, Vol. 17, pp. 87-97, 98.

[11] M. Grossberg, H. Peri, S. Nayar, P. Belhumeur, Making One Object Look Like Another, CVPR, pp. 452-459, 04.

[12] S.Nayar, V.Branzoi, T.Boult, Programmable Imaging using a Digital Micromirror Array, CVPR, pp.436-443, 04.

[13] P.Debevec, J.Malik, Recovering high dynamic range radiance maps from photographs, SIGGRAPH, pp. 369-378, 97.

[14] M.Robertson, S. Borman, R. Stevenson, Dynamic range improvement through multiple exposures, ICIP, pp. 159-163, 99.

[15] S. Kang, M. Uyttendaele, S. Winder, R. Szeliski, High dynamic range video, SIGGRAPH, pp. 319-325, 03.

[16] Y.Boykov, V. Kolmogorov, What Energy Functions Can Be Minimized via Graph Cuts?, PAMI, vol. 26, nr. 2, pp. 147-159, 04.