Cavlectometry: Towards Holistic Reconstruction of Large Mirror Objects

Jonathan Balzer
UCLA
balzer@cs.ucla.edu

Daniel Acevedo-Feliz
KAUST
daniel.acevedo@kaust.edu.sa

Stefano Soatto
UCLA
soatto@cs.ucla.edu

Sebastian Höfer
KIT
sebastian.hoefer@ies.uni-karlsruhe.de

Markus Hadwiger
KAUST
markus.hadwiger@kaust.edu.sa

Jürgen Beyerer
Fraunhofer IOSB
juergen.beyerer@iosb.fraunhofer.de

Abstract

We introduce a method based on the deflectometry principle for the reconstruction of specular objects exhibiting significant size and geometric complexity. A key feature of our approach is the deployment of an Automatic Virtual Environment (CAVE) as pattern generator. To unfold the full power of this experimental setup, an optical encoding scheme is developed which accounts for the distinctive topology of the CAVE. Furthermore, we devise an algorithm for detecting the object of interest in raw deflectometric images. The segmented foreground is used for single-view reconstruction, the background for estimation of the camera pose, necessary for calibrating the sensor system. Experiments suggest a significant gain of coverage in single measurements compared to previous methods.

1. Introduction

1.1. Motivation

Reconstructing the geometry of an object or scene from a collection of images is an important problem in applications ranging from assisted surgery to additive manufacturing, quality control, augmented reality, just to mention a few. The problem is particularly challenging when objects exhibit complex non-Lambertian reflectance. For instance, specular surfaces cannot be reconstructed without some assumption on the surrounding environment. While generic prior assumptions may be sufficient for classification purposes, accurate metrology calls for precise knowledge of the surroundings, that can be thought of as the “illumination”. In deflectometry, a controllable illuminant (e.g., a liquid-crystal display, LCD) is reflected through the object’s surface onto the image plane, see Fig. 2(a). In order to reconstruct a model of the unknown surface from the known illuminant (input) and its measured image output, the input must be sufficiently exciting. Typically, a series of optical code words is displayed, identifying each of the points on the illuminant uniquely. This way, dense correspondence between the domains of input and output signal is established, the so-called light map, assigning to a pixel the scene point it sees via the specular surface [10].

To make the basic deflectometry procedure viable for industrial applications, an important open problem needs to be solved: Convex objects, no matter how small, act as demagnifying lenses and thus reflect a significant portion of the environment. For instance, in Fig. 1(a), almost half of

The terms scene, illuminant, and pattern generator are used synonymously throughout this paper.

Figure 1. Objects of study in some of the current literature on mirror surface reconstruction [1, 20, 38, 3]: (a) Convex surface patches compress the reflected image of the environment. (b) - (c) This effect can be remedied by placing the surface close to the pattern generator and sampling a large number of vantage points. (d) The markings on this engine hood give an impression of how many measurements were required by the method described in [3].
the scene becomes visible in a spherical patch of just a few square centimeters. There are essentially two countermeasures: One can either decompose the surface into smaller tractable pieces, cf. [3], but depending on the geometry of the object, the number of required views can easily reach the hundreds. Alternatively, many state-of-the-art methods optimize the effective area on the mirror surface by placing the pattern generator in close proximity, cf. Fig. 1(b). Going one step further, Hong et al. [15] arrange five LCDs tightly around the object leaving a small gap in between for a single camera to peek through. Both strategies, however, make it difficult if not impossible to account for occlusions, the crux of any reconstruction methodology. In fact, one of the idiosyncrasies of deflectometry is that it may suffer from occlusions not just of the surface itself but also of the illuminant, i.e., light rays can be blocked after interacting with the surface and before. For the purpose of reconstruction, each point on the surface must grant at least one unobstructed view on the illuminant, and ideally, images should be acquired from different camera positions until this condition is completely satisfied. To this end, Weinmann et al. [38] furnish their system with some of the desired flexibility by combining eleven cameras and three LCDs with a turntable to carry the specimen. While their approach yields impressive reconstructions of small-scale objects, first, it still relies on a very large number of measurements, 792 for the example in Fig. 1(c), and second, it excludes large surfaces, e.g., automobile parts.

1.2. Contribution and overview

Here, we propose not only to extend the illuminant to comprise the entire scene, but to do this in a way that supports acquisition of objects more than a meter in length from multiple vantage points. More concretely, we prove the feasibility of performing deflectometry in an Automatic Virtual Environment (CAVE), which we dub cavlectometry. A CAVE is a room-sized cube consisting of up to six back-projected walls, see Figs. 2(b) and (c). In its default operating mode, users stand inside the CAVE and, while wearing stereo glasses and tracking devices, experience virtual 3-d scenes [8]. Taking advantage of it as an illuminant for deflectometry sparks a series of innovations: Rather than treating each of the walls separately, the encoding scheme introduced in Sect. 2.1 – our first contribution – takes into account the special properties of the CAVE. In Sect. 3, we develop a calibration method which requires no additional data besides that used in later reconstruction, at the heart of which lies an algorithm for separating image regions arising from either Lambertian or specular surfaces. This is our second contribution. Cavlectometry allows us to reconstruct surface patches of unprecedented size from a single vantage point (Sect. 5.2). Public deflectometry benchmarks are scarce with [3] being the only notable exception. Thus, the final contribution we offer is to distribute all of our data on the web site of the first author. Before we explain our method in detail, let us next put it in the context of the existing literature on specular surface reconstruction.

1.3. Other related work

Normals of surfaces with hybrid reflection properties can be recovered from purely radiometric considerations [17, 13, 31]. Not before correspondences are available between pixels and the scene points they portray, tools from geometric optics can be leveraged: identifying a single point light source renders the computation of the light map trivial; then, under known camera motion, an initial point can be expanded into a surface curve by tracking the location where the highlight first appeared [41]. The standard structure-from-motion pipeline can also be enriched as to explicitly take into account specular reflections of a discrete set of scene points [32]. Zheng and Murata [40] progress from isolated point sources towards a one-dimensional concentric illuminant. The accuracy of their approach however remains limited as long as some points on the light source remain indistinguishable. Circular reflection lines suffice nevertheless for special applications such as measuring the cornea in human eyes [12] or surface interrogation [9, 33]. Savarese and Perona [30] study reflections of a two-dimensional checkerboard pattern.
Controllable illuminants were introduced to boost reliability and density of correspondences: The very first active deflectometry setup – due to Sanderson et al. [29] – contained an array of light-emitting diodes (LEDs) which could be switched on sequentially to detect them in the camera image. The acquisition time can be reduced by showing binary codes in parallel [23]. Both papers address the ambiguity by assuming quasi-parallel illumination. A large body of literature on the subject exists in the field of optical metrology starting with [25] that suggests improvements if the LED array is replaced by a commodity computer monitor. Knauer et al. [18] describe the theoretical limits of the method involving phase-shifted sine patterns as codes. Light map measurements can be further enhanced by color displays [35]. An active variant of Savarese’s method is presented in [28].

Gauge ambiguities can be eliminated by integrability considerations [14, 20] or specular stereo [37, 10, 6, 38]. In the special case of infinitesimal displacement of any participant of the imaging chain – camera, surface, or scene – the deflectometric image is perturbed by a specular flow field. Blake treats the case of a moving observer in a series of papers culminating in [5], which Waldon and Dyer [36] extend by adding photometric effects and interreflections to the model. Like in [41], the trajectories of virtual features in the image plane give rise to reconstructions of one-dimensional subsets of the sought-after surface [24]. They are linked with the caustic curves known in geometric optics [34]. Roth’s and Black’s [27] method delivers dense two-dimensional reconstructions from specular flow generated by a known camera motion. The authors utilize theoretical results from [7] where, with the objective of speeding up ray tracing along piecewise linear paths, it is shown how the points of deflection, i.e., the vertices of the path, behave w.r.t. to changes of the ray emitter. Shape cues may be recovered from the flow field alone, i.e., without image-scene correspondences, when the scene is moving relative to object and camera [1]. A closed-form model of specular flow on arbitrarily deforming surfaces is derived in [19]. Estimation of specular flow hinges only on the natural scene and can thus handle surfaces of arbitrary size and shape complexity but, as a downside, it does not permit metrically precise reconstruction. A comprehensive summary of the state of the art in deflectometry can be found in several recent survey papers [16, 4, 26].

2. Data acquisition

2.1. Measurement of the light map

In the following, matrices are in bold, vectors bold italic. Vectors $\mathbf{x} \in \mathbb{R}^3$ of unit length are marked by a hat, i.e., $\|\hat{\mathbf{x}}\| = 1$. Each pixel in the image plane of a calibrated camera corresponds to a direction $\hat{\mathbf{x}}$, under which the associated sight ray leaves the projection center. The relationship between the two is determined by the intrinsic parameters of the camera. Where the ray intersects the surface in $\mathbf{x}$, it gets “deflected” according to the geometric law of reflection and travels on until it hits a scene point $\mathbf{l}$. For the purposes of geometric modeling, the CAVE is a cube $C$ with boundary $\partial C$ and side length $2h = 3048$ mm. If we neglect multiple reflections and occlusions induced by the camera, the scene point comes to lie on one of the faces of $C$. Apart from the object under inspection, the scene is $\partial C$ and vice-versa. Altogether we obtain a mapping $\mathbf{l} : S^2 \to \partial C$, which assigns to a viewing direction $\hat{\mathbf{x}}$ the point $\mathbf{l}$ seen by the pixel associated with that direction, see Fig. 3. This is precisely the light map defined in Sect. 1.1. Its measurement must be the first step of any deflectometric method.

2.1.1 Phase shifting

Although optical phase-shift codes originate from a variety of established applications like optical interferometry, synthetic aperture radar, and magnetic resonance imaging, its features make it the ideal candidate for estimation of the light map: Gradually phase-shifted sine patterns facilitate subpixel accuracy and are relatively insensitive to disturbances of absolute intensity values, e.g., caused by staining of the specular surface. As in a typical setup (Fig. 2) the camera is focused onto the surface under inspection, limited depth of field exposes the imaged code patterns to blurring. A sine pattern, however, can pass the optical system – which in this case resembles a low-pass filter – unaffected as long its spatial frequency is sufficiently low.

Three phase shifts are sufficient to recover a single spatial coordinate uniquely; additional ones may compensate for non-uniformities (e.g., from gamma nonlinearity, etc.). Owing to the fact that phase information is uniquely defined only within the range of one wavelength of the pattern, decoding must be complemented by a so-called phase unwrapping. Explaining the codification methodology in detail is beyond the scope of this paper; we refer the interested
reader to [11] among others. We do emphasize here that care must be taken when displaying phase-shifted sine patterns on the walls of the CAVE simultaneously, an issue that never arose in previous systems which were all designed to run on disjoint rectangular screens. The naive approach would be unfolding the cube into the canonical net for cube mapping and shift the patterns along the axes of some local coordinate system. Experimental evidence, however, suggests that treating each face regardless of the others causes problems at the transitions between two maps, see Fig. 4. Here, due to self-illumination, a pattern leaks from one wall into the other. Our key insight is that this is uncritical so long two patterns are symmetric across an edge of ∂C, i.e., if they do not meet orthogonally, at different frequencies or phases. An encoding scheme which fulfills this requirement by moving patterns only parallel to the edges of ∂C is shown in Fig. 5.

2.1.2 Face encoding

The inherent structure of this scheme is easily verified: each subsequence encodes one of the three global Cartesian coordinates. Only a single bit of information is missing, prompting a second coding stage. To illustrate the problem, suppose we are looking for the location of a point l on one of the CAVE walls, say \( l = (l_x, l_y, l_z) \) where \( l_x = h \). Applying the sequences in Fig. 5 will provide the values of \( l_y \) and \( l_z \) there. The subsequence shifting in \( x \)-direction will leave \( l \) dark. Putting these facts together, we conclude that \( l \) will be on a plane \( x = \text{const} \), but we do not know whether it is the one defined by the equality \( x = h \) or \( x = -h \). The sign of \( h \) is exactly the missing bit. A single image should suffice to set it correctly. The problem of self-illumination, however, persists. In fact, the leakage of a pattern shifting in one of the three coordinate directions into the two orthogonal faces that remain dark meanwhile is even more severe. For this reason, we resort to a complete detection of each of the faces of \( C \). There are altogether six prompts prompting the use of a binary code of length 3 bit. An illuminated wall represents a bit value of 1, see Fig. 6. Robustness is further improved with the help of a differential code: Thereby, a bit value of 1 is encoded by the transition from white to black. Conversely, a bit value of 0 corresponds to an on-switching of the illumination. The differential face coding amounts to in total six images per light map measurement.

3. Calibration

In order to avoid screen occlusions, we advocate recording multiview datasets in a monocular fashion, whereby measurements are taken sequentially while the camera is displaced between two captures. To convert a measured correspondence between \( \hat{x} \) and \( l = l(x) \) into a reasonable normal estimate, both vectors have to be represented in a common reference frame. On one hand, \( \hat{x} \) is usually given w.r.t. the frame that is attached to the camera and moves along with it to each new vantage point. On the other hand, \( l \) is naturally expressed in terms of the frame at the barycenter of the CAVE (Fig. 5(a) top), which remains static and is thus selected as our world coordinate system. The transformation between the two is an element \( g = (R, t) \), \( R \in \text{SO}(3), t \in \mathbb{R}^3 \), of the rigid-motion group \( \text{SE}(3) \), which – as initially unknown – needs to be determined through calibration.

It turns out that \( g \) comes as a byproduct of the very same light map \( l \) used for reconstruction later on (Sect. 4). The basic idea of the following procedure is to identify those
regions that are imaged without intermediate reflections and
to localize the camera by means of the appendant scene-
image correspondences:

1. Consider the pullback of the light map \( l^* : D \to \partial C \)
to the image plane \( D \subset \mathbb{R}^2 \). It can be calculated
knowing the intrinsic parameters of the camera, which
we initially estimate by a standard method \([39]\). The
extrinsic calibration process starts with a – possibly
very coarse – segmentation of \( l^* \), which builds upon
the fact that walls seen directly generate piecewise
homographies. One of the distinct properties of ho-
mographies is that they maintain cross-ratios, whose
computation, however, may be numerically unstable.
Instead, we propose a simpler – less discriminative
yet effective – inference criterion based on colinear-
ity. Denote by \( l_{ij} \) the value of the light map at pixel \( p := (u_i, v_j) \in D \) in the discretized image plane. If
\( l^* \) behaves locally like a homography around \( p \), then
the points \( l_{i-1,j}, l_{i,j}, l_{i+1,j} \) and \( l_{i-1,j}, l_{i,j}, l_{i,j+1} \)
will be respectively colinear, in other words, \( r_{ij}^x = r_{ij}^y = 1 \)
where

\[
\begin{align*}
    r_{ij}^x := \text{rank}(l_{i-1,j}, l_{i,j}, l_{i+1,j}) \\
    \text{and} \\
    r_{ij}^y := \text{rank}(l_{i,j-1}, l_{i,j}, l_{i,j+1}).
\end{align*}
\]

We can define a scoring function \( s : D \to \mathbb{N} \) by

\[
s(u_i, v_j) := |1 - r_{ij}^x r_{ij}^y|,
\]

which vanishes where both rank conditions are satis-
ified (Fig. 7(a)). A first segmentation of \( l \) respectiv-
ely \( l^* \) is obtained by binarizing \( s \) with threshold 0.5
(Fig. 7(b)). Following morphological enhancement,
this segmentation is refined with the help of a wasser-
shed transform \([22]\), see Fig. 7(c).

2. Observe that there are degenerate cases in which the
specular surface itself affords a homography, e.g., if it
is locally planar or reflects into a single point (i.e., the
light map is constant). We can, however, safely assume
that the background occupies a significant amount of
the image. Otherwise, we would hardly ever get access
to \( g \) at all. With this in mind, we shrink the background
region determined in the previous step to those pixels
that map to the wall which appears most often in it.

3. The homography between the image plane and the re-
sulting background region, i.e., the CA VE wall that ap-
ppears in most pixels and is seen directly, is computed
by a standard direct linear transformation method.
Given the projection matrix, \( g \in SE(3) \) can be ex-
tacted from the associated homogeneous matrix \([39]\).

4. We now revisit the full light map \( l^* \) on the image plane.
Every point \( l(p) \) that can be projected directly into the
image plane, utilizing the initial estimate of \( g \), must
belong to the background. The characteristic function
of the foreground then becomes

\[
\chi_b(p) := \begin{cases} 
1 & \text{if } \|\pi(RL(p) + t) - p\| < \theta, \\
0 & \text{otherwise},
\end{cases}
\]

where \( \pi \) denotes the canonical pinhole projection. Ob-
serve that the sensitivity of \( \chi_b \) w.r.t. the threshold is
quite low. In other words, the choice of \( \theta \) requires no
great care because the backprojection error grows dis-
continuously across the silhouette. The final result is
depicted in Fig. 7(d).

---

Figure 6. (a) A binary code identifies each separate wall uniquely.
(b) Visualization of the full vector-valued light map \( l \). The color
map assigns an HSV coordinate to each \( l \in S^2 \) with hue and
saturation varying proportionally to azimuth respectively elevation
while the value is kept constant at 1.0.

Figure 7. Segmentation of deflectometric images of strongly-
specular objects: (a) Foreground score based on local ranks of the
light map. (b) Initial segmentation of (a) by thresholding. (c) En-
hancement by watershed transform. (d) The final segmentation is
obtained after extrinsic calibration by thresholding the backprojec-
tion error.
With the final segmentation $\chi_b$, we improve the estimate of the extrinsic camera parameters $g$ by running a robust bundle adjustment on all points in the background [39]. Let us remark that in previous experimental setups, the screen is rarely seen directly during measurement but obviously has to be for extrinsic calibration. To establish an oblique view on the screen, a calibration normal (usually a planar mirror) is necessary, which complicates Step 3 and the final bundle adjustment significantly [18]. Also note that the procedure described in [21] is not applicable here as it requires the pattern generator to be movable.

4. Surface reconstruction

What sets the deflectometric measurement principle apart from techniques such as stereopsis or laser scanning is that it yields slope which must be converted into depth by means of integration. More precisely, the light map $l$ acquired from a vantage point $g$ is equivalent to a unit vector field of measured normals $\hat{n}_m(x)$ in space. To see this, fix a point $x \in C$ inside the visible volume, transform it to camera coordinates via $g^{-1}x$, project it to $S^2$ by normalization, and look up the value of the light map $l = l(\hat{x})$ (Fig. 3). Suppose that $x \in S$, then the surface normal $\hat{n}(x)$ must obey the law of reflection, i.e., the correspondence between $x$ and $l$ yields a hypothesis $\hat{n}_m$ about the normal the true surface should have if it indeed contained the point $x$. We say that a regular surface $S$ solves the reconstruction problem if it interpolates the measured normal field in the sense that $\hat{n}(S) = \hat{n}_m|_S$ except on zero-measure subsets of $S$. Such an $S$ rarely exists due to stochastic disturbances of the data. A maximum-likelihood estimate of the surface is obtain as the solution of the following variational problem:

$$S^* = \arg \min_S \int_C \frac{1}{2} \|\hat{n}(x) - \hat{n}_m(x)\|^2 \, dx. \quad (2)$$

The well-known gauge ambiguity manifests itself in the dependence of $\hat{n}_m$ on $x$: an additional constraint $x_0 \in S$ needs to be imposed to render problem (2) well-posed [4].

Since we are mainly concerned with the deflectometric data acquisition process and less with reconstruction, we assume the point $x_0$ to be known for now. The next section will outline a suitable initialization heuristic. For more sophisticated regularization approaches, we refer the interested reader to the related literature, see Sect. 1.3. Also, at this point, we make no attempt to fuse data acquired from a series of vantage points into a common normal field. Our data could be processed with any existing multiview normal field integration scheme, cf. [6, 38].

5. Experimental evaluation

5.1. Implementation

We implemented most of the pipeline in C++, except for the steps related to optical coding. All experiments were run on a single 3.4 GHz core of a commodity computer. We solve the shape optimization problem (2) with the algorithm proposed in [2]. As an initial guess, we use a fronto-parallel plane which is also tangent to the visual hull. Loop subdivision is applied to the evolving mesh between two steps to increase the level of detail and prevent the iteration from falling into local minima of the cost functional\(^2\), see Fig. 10.

5.2. Results

A selection of segmentation masks obtained with the method developed in Section 3 is shown in Fig. 8. The computation of a single mask constitutes the main bottleneck

\(^2\) although this is less critical given that the method is of second order

Figure 8. Segmentation masks. Pixels with invalid measurements are shown in gray (e.g., on Lambertian parts of the object, the edges of the CAVE, or where the camera occludes the walls).

Figure 9. Ratio of the area covered by projections of surface points on the image plane (in percentage of pixels) over the area on the screen (parametrized by the distance $d$ to a fixed point) which is seen through these points. Focal length: 16.3 mm.

Figure 10. Four iterations of the normal integration method with intermediate uniform refinements of the triangular mesh.
Figure 11. A picture of the specimen and our reconstruction colored by local deviation from reference model in mm.

Figure 11. A picture of the specimen and our reconstruction colored by local deviation from reference model in mm.

Table 1. Performance statistics.

<table>
<thead>
<tr>
<th>Object</th>
<th>No. faces</th>
<th>No. views</th>
<th>Area [m²]</th>
<th>Mean error [mm]</th>
<th>Max. error [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunny</td>
<td>26681</td>
<td>1</td>
<td>0.57</td>
<td>0.002</td>
<td>0.594</td>
</tr>
<tr>
<td>Moka pot</td>
<td>25630</td>
<td>1</td>
<td>0.015</td>
<td>0.27</td>
<td>3.69</td>
</tr>
<tr>
<td>Teapot</td>
<td>61922</td>
<td>1</td>
<td>0.011</td>
<td>0.39</td>
<td>2.8</td>
</tr>
<tr>
<td>Hood</td>
<td>44588</td>
<td>1</td>
<td>0.76</td>
<td>1.12</td>
<td>13.6</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, we have argued that deflectometric imaging and reconstruction benefit strongly from illuminants that encompass all of the scene except for the mirror object under inspection. These can be realized e.g. by means of a CAVE. A low-cost alternative in a regular room with opaque white walls and cheap standard projectors is shown in Fig. 2(d). In conclusion, let us remark that the CAVE, beyond its role as deflectometry illuminant, may be utilized for reconstruction in general.
Figure 12. Closing the loop: real images acquired in the CAVE (top) compared to simulations (bottom) based on vantage points and object models obtained by our method.

Acknowledgements

This work was supported by ONR N00014-13-1-0563 and ARO W911NF-11-1-0391.

References