Appearance and Shape from Water Reflection

Ryo Kawahara Meng-Yu Jennifer Kuo Shohei Nobuhara Ko Nishino
Kyoto University, Kyoto, Japan
http://vision.ist.i.kyoto-u.ac.jp/

Abstract

This paper introduces single-image geometric and appearance reconstruction from water reflection photography, i.e.*, images capturing direct and water-reflected real-world scenes. Water reflection offers an additional viewpoint to the direct sight, collectively forming a stereo pair. The water-reflected scene, however, includes internally scattered and reflected environmental illumination in addition to the scene radiance, which precludes direct stereo matching. We derive a principled iterative method that disentangles this scene radiometry and geometry for reconstructing 3D scene structure as well as its high-dynamic range appearance. In the presence of waves, we simultaneously recover the wave geometry as surface normal perturbations of the water surface. Most important, we show that the water reflection enables calibration of the camera. In other words, for the first time, we show that capturing a direct and water-reflected scene in a single exposure forms a self-calibrating HDR catadioptric stereo camera. We demonstrate our method on a number of images taken in the wild. The results demonstrate a new means for leveraging this accidental catadioptric camera.*

1. Introduction

Water reflection has long been a source of artistic inspiration. Various paintings come to mind that compose reflection by a water surface together with direct sight of a scene, such as Claude Monet's Autumn in Argenteuil. Water reflection has also been an integral part of architectural design as seen in Taj Mahal and Matsumoto Castle to name a few. Water reflection has also been used as an artistic expression in modern photography, for instance, by capturing a cityscape reflected in a puddle.

It is perhaps much less understood that water reflection carries visual cues for scene structure recovery. A computer vision researcher, however, would likely notice that water reflection would give a different vantage point of the scene from the camera viewpoint when captured in a single image, collectively forming a (flipped) stereo pair. This suggests an opportunity for single-image scene geometry recovery. In fact, Yang et al. [32] applied standard stereo reconstruction to estimate scene depth from a single water reflection image after adapting reflected scene appearance to construct a cost volume robust to their radiometric distortions.

Figure 1. We show that we can recover the 3D geometry (right) and high-dynamic range appearance (middle) of a scene from a single image (left) capturing it both directly and through reflection by a water surface.

In this paper, for the first time to our knowledge, we show that a single image capturing both the direct and reflected observation through water reflection of a scene results from a self-calibrating high-dynamic range catadioptric imaging system. That is, in sharp contrast to merely leveraging the geometric configuration of water reflection, we show that a high-dynamic range appearance and 3D shape of the scene can be recovered without any knowledge about the image formation a priori. We first consider the case where the water surface is calm and can be modeled as a planar mirror. As shown in Fig. 1, we derive a method that recovers high-dynamic range appearance and 3D structure of the scene. The main challenge lies in the fact that water-reflected scene radiance is compounded with environmental light scattered in the water medium and also reflected by the bottom surface. Scene radiance must be sifted out from this superposition in order to match against the direct observation for triangulation. In other words, radiometry and geometry recovery are inherently intertwined.

We derive a canonical iterative method to recover scene geometry from the direct and water-reflected observations. We also show that high-dynamic range scene radiance can be estimated in the process and water reflection even enables calibration of the camera. That is, we do not need to know anything about the camera; its intrinsic parame-

ters can be recovered by seeking agreements in angular-dependent Fresnel effects in the reflected observation, and its extrinsic parameters can be estimated by identifying the water surface. In other words, we show that capturing direct and water-reflected scenes in a single exposure forms a self-calibrating HDR catadioptric stereo camera.

The water surface in the image is not always calm and can have waves that lead to noticeable displacements in the reflected observation. We show that this can be modeled as surface normal variations of the water surface and derive an iterative approach to simultaneously recover the shape of both the scene structure and the water surface. For this, we introduce a principled method for incorporating realistic prior knowledge such as piecewise planarity for scene geometry and a Fourier-domain wave representation.

We experimentally validate our method both quantitatively and qualitatively by reconstructing scene structure and radiance from a wide range of real images. The results agree with the newly derived theory and demonstrate the effectiveness of its application to arbitrary images taken in the wild. The proposed method provides a new means for visually appreciating our 3D world, and enables a new form of 2D-3D visual media of *water reflection photography*.

2. Related works

To our knowledge, the work by Yang et al. [32] is the only other work that recovers 3D from an image of water reflection. This work modifies a standard stereo algorithm to compute two cost volumes, one for direct and another for reflected views, whose filtered disparities are later fused to produce a depth map. The method uses automatically established keypoint pairs between the direct and reflected views to adapt the appearance and also limit the range of disparities. This inevitably necessitates the automatically detected keypoints to uniformly span both the spatial and depth variations of the scene, which hinders the method's applicability–their results are all on well-textured natural scenes. The radiometric distortions in the reflected view are assumed to be rectifiable with simple linear adaptation both locally and globally. This assumption is simply incorrect due to the compound angular-dependent mixture of light as we later show and model. In contrast, instead of treating them as nuisance that needs to be corrected, we exploit the unique radiometric properties of water reflection as a rich source for true scene appearance recovery and show that it also enables self-calibration of the camera. That is, we show how to recover not just the geometry but also the radiometry of a scene, which are inherently intertwined, from a single water reflection image.

Appearance and shape from water reflection can be interpreted as accidental catadioptric stereo imaging in which the water surface serves as the catoptric view that forms a stereo pair with the direct dioptric view. Examples of accidental imaging include the use of occluders as pinspeck (anti-pinhole) cameras that capture surroundings as shadows [28], which can be used to estimate the scene behind the occluders [5, 4]. Accidental micro motions due to, for example, heart beating, can provide scene depth cues that can be used for image refocusing and synthetic parallax generation [33, 10, 9].

Catadioptric imaging [2, 25] in computer vision has been applied to a variety of tasks including omnidirectional imaging [22], reflectance acquisition [16], shape-from-silhouette [7], structured light [13, 26], kaleidoscopic imaging [20, 27], and stereo reconstruction [1, 17, 24]. Gluckman and Nayer [8], in particular, proposed a catadioptric stereo system with two planar mirrors. As the first example of an accidental catadioptric system, Nishino and Nayar [18, 19] showed that capturing eyes form a catadioptric imaging system in which the cornea serves as the reflector.

Stereo matching with translucency [29, 31] or image-based reflection separation [14, 21, 12, 23] explicitly model transmission through semi-translucent surfaces. They utilize either 3D recovery or models in the Fourier domain for blind separation of reflected and transmitted images. They cannot, however, be applied to non-planar surfaces such as wavy water surfaces.

3. Assumptions

Let us first clarify the assumptions we make. As we consider an image capturing both a direct view of a scene and its reflection by a water surface, we can safely make the following assumptions without loss of generality.

- The water medium (*e.g.*, pond or puddle) is homogeneous and has a known index of refraction.
- The reflected scene as well as the bottom of the water medium consist of Lambertian surfaces.
- The reflection is purely specular at the water surface.
- The sun is not captured in the reflection.
- We can either manually or automatically isolate the image region capturing water reflection.

We do not require knowledge of the camera parameters neither intrinsic nor extrinsic. If EXIF information is available, we use it to initialize the intrinsic camera parameters. We show that these camera parameters can be estimated from the image.

When the water surface has waves, we assume that they satisfy the following properties.

- The wave amplitude (*i.e.*, height) is comparatively smaller than the camera height from the water surface.
- Any interreflection on the water surface is negligible in intensity and one water surface point reflects one (but not necessarily unique) object surface point.

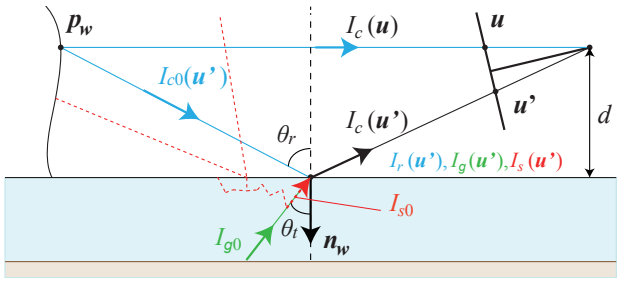

Figure 2. A scene point $\boldsymbol{p}_w$ is observed directly ($\boldsymbol{u}$) and also through reflection $\boldsymbol{u}'$ by the water surface. Unlike the direct observation $I_c(\boldsymbol{u})$ the reflected radiance $I_c(\boldsymbol{u}')$ contains not just the direct radiance from the scene point $I_{c0}(\boldsymbol{u}')$ which gets specularly reflected by the water surface, but also the radiance of light that has scattering through the water medium I_{s0} and that reflected off the bottom surface of it I_{g0} that is refracted into the camera.

4. Planar Water Reflection

We begin by deriving the key steps for recovering appearance and shape from planar water reflection which are also shared when dealing with wavy water reflection. As depicted in Fig. 2, a 3D scene point $\boldsymbol{p}_w$ is observed twice in the image: the direct observation $I_c(\boldsymbol{u})$ and the reflected observation $I_c(\boldsymbol{u}')$. Note that both *observations* of scene points are captured in a single image and represented by their positions on the image plane, $\boldsymbol{u} = (u_x, U_y, 1)^\top$ and $\boldsymbol{u}' = (u'_x, U'_y, 1)^\top$, respectively.

4.1. Planar Water Surface Reconstruction

For a calm water surface, or as an initial estimate of the global surface normal of a wavy water surface, we estimate the surface normal of the water surface $\boldsymbol{n}_w$ from a small set of corresponding pairs of direct and reflected observations $(\boldsymbol{u}, \boldsymbol{u}')$ satisfying

$$\boldsymbol{u}'^\top A^{-\top} [\boldsymbol{n}_w]_\times A^{-1} \boldsymbol{u}' = 0\,, \tag{1}$$

where A is the intrinsic parameter matrix of the camera and $[\boldsymbol{n}_w]_\times$ is the skew-symmetric matrix of $\boldsymbol{n}_w$. This is a linear constraint on the normal $\boldsymbol{n}_w$, and we can estimate $\boldsymbol{n}_w$ from two or more corresponding pairs. Intrinsic camera parameters A can be initialized with EXIF information, when available, or with reasonable values common in outdoor photography, which are then refined by the radiometric recovery process as we detail in Sec. 5.5.

We obtain these correspondences semi-automatically. We first segment the image into direct and reflected observation regions, which can often be done by just specifying the line where the direct and reflected observations meet in the image. We then run generic feature detection and matching methods in these two regions. For a calm water surface, we found this process to be sufficient for all cases. For a moderately wavy water surface, we conduct this automatic feature matching on a downsampled and blurred image to obtain the global surface normal of the water surface. Note that we only need a few correspondences as we are only recovering the water surface normal.

4.2. Direct–Reflected Stereo Reconstruction

For a calm water surface that can be modeled as a planar reflector, once we estimate its normal, the direct and reflected observations in the image form a stereo image pair (albeit folded). For a wavy water surface, the stereo correspondence pairs are locally perturbed by the varying surface normal at each water surface point. In either case, if the displacements due to waves are undone and correspondences are established as we later show, we may recover the 3D coordinates of scene points via regular stereo reconstruction (*i.e.*, triangulation) from the direct–reflected observation point pairs:

$$\begin{cases} \boldsymbol{u} &= \lambda_c A \boldsymbol{p}_w\,, \\ \boldsymbol{u}' &= \lambda'_c A (H_w \boldsymbol{p}_w + \boldsymbol{t}_w) \end{cases} \Leftrightarrow \begin{pmatrix} u_x M_3 - M_1 \\ u_y M_3 - M_2 \\ u'_x M'_3 - M'_1 \\ u'_y M'_3 - M'_2 \end{pmatrix} \boldsymbol{p}_w = 0, \tag{2}$$

where $H_w = (I - 2\boldsymbol{n}_w \boldsymbol{n}_w^\top)$ is a householder matrix, and $\boldsymbol{t}_w = 2d\boldsymbol{n}_w$. $M = (A\ \mathbf{0})$, $M' = A(H_w\ \boldsymbol{t}_w)$ are projection matrices for each viewpoint.

This stereo reconstruction requires the intrinsic and extrinsic parameters of the camera. The extrinsic camera parameters can be described by the mirrored camera pose H and translation $\boldsymbol{t}_w$ w.r.t. the original viewpoint defined by the water surface normal $\boldsymbol{n}_w$ and distance d of the camera from the water surface as depicted in Fig. 2. Since the global scale is not known, we assume $d = 1$ and the recovered scene geometry is scaled accordingly.

5. Appearance from Water Reflection

As we show in Fig. 2, the reflected observation of a scene point $I_c(\boldsymbol{u}')$ is a superposition of specular reflection by the water surface $I_r(\boldsymbol{u}')$ of the Lambertian scene radiance $I_{c0}(\boldsymbol{u}')$, environmental light scattered through the water medium $I_s(\boldsymbol{u}')$, and also bouncing off the bottom surface $I_g(\boldsymbol{u}')$

$$\begin{aligned} I_c(\boldsymbol{u}) &= I_{c0}(\boldsymbol{u}')\,, \\ I_c(\boldsymbol{u}') &= I_r(\boldsymbol{u}') + I_g(\boldsymbol{u}') + I_s(\boldsymbol{u}')\,. \end{aligned} \tag{3}$$

We need to separate these components and recover the scene radiance $I_{c0}(\boldsymbol{u}')$, so that correspondences can be established between the reflected and direct observations.

5.1. Reflected Scene Radiance

The reflected scene radiance by the water surface I_r can be described by the Fresnel power reflectance F

$$I_r(\boldsymbol{u}') = F(\boldsymbol{u}')I_{c0}(\boldsymbol{u}')\,. \tag{4}$$

The power reflectance F under natural light is

$$F(\boldsymbol{u}') = \frac{1}{2}(R_s + R_p)\,, \tag{5}$$

where R_s and R_p are the power reflection coefficients for s-polarized and p-polarized light given respectively by

$$R_s = \left(\frac{\sin(\theta_r - \theta_t)}{\sin(\theta_r + \theta_t)}\right)^2\,, \quad R_p = \left(\frac{\tan(\theta_r - \theta_t)}{\tan(\theta_r + \theta_t)}\right)^2\,. \tag{6}$$

The angle of incidence θ_r, which is also the angle of specular reflection, is

$$\theta_r = \cos^{-1}\left((A^{-1}\boldsymbol{u}')^\top \boldsymbol{n}_w\right)\,, \tag{7}$$

and the angle of refraction θ_t becomes

$$\theta_t = \sin^{-1}\left(\frac{\mu_a}{\mu_w}\sin(\theta_r)\right)\,, \tag{8}$$

from Snell's law. μ_a, μ_w are the refraction indices of the air and water respectively.

5.2. Water-Scattered Environmental Illumination

We assume that the environmental illumination is uniform across the water surface, *i.e.*, the incident environmental illumination to the water surface points in the reflected-observation region does not vary. This is a reasonable assumption as long as the sun is not directly imaged via water reflection. Part of this environmental light is transmitted into water, scattered in the medium, and then transmitted back into the viewpoint.

The scattered environmental illumination is the sum of scattered light from all points along the transmitted path observed at $\boldsymbol{u}'$. On the other hand, the bottom of the water medium at which the light path intersects will have a comparatively weaker radiance, which suggests that we can limit the contribution of water-scattered environmental illumination to that from near the water surface. We model this near-surface water-scattered environmental illumination based on the dipole method [11], which describes the process of environmental illumination $I(\theta_i)$ from θ_i transmitted into water with $T(\theta_i) = 1 - F(\theta_i)$, attenuated with $R_d(\tau)$, then transmitted again into the camera $T(\theta_r(\boldsymbol{u}'))$ for all angle θ_i and distance τ:

$$\begin{aligned} I_s(\boldsymbol{u}') &= \int_\tau \int_{\theta_i} T(\theta_i)R_d(\tau)T(\theta_r(\boldsymbol{u}'))I(\theta_i)d\tau d\theta_i\,, \\ &\simeq T(\theta_r(\boldsymbol{u}'))I_{s0}\,, \end{aligned} \tag{9}$$

Figure 3. The Fresnel reflection on the water surface, in effect, provides scene radiance captured with spatially varying exposures that are different from the camera. By combining these two radiance exposures, we reconstruct high-dynamic range appearance of the scene (right), which reveals scene details in saturated regions in the original direct view (left).

where I_{s0} is the homogeneous scattered radiance. Note that, for water, the BSSRDF is simply $T(\theta_i)R_d(\tau)T(\theta_r(\boldsymbol{u}'))$, since the scattering attenuation only depends on the distance τ between the incident and exitant surface points.

Note that, when the medium is shallow, the bottom surface reflection will be comparatively strong and instead be transmitted into the viewing direction almost without any scattering. This effect will be accounted for by the bottom surface reflection model described in Sec. 5.3.

5.3. Reflection From The Bottom Surface

The reflected observation also includes environmental illumination reflected by the bottom surface of the water medium. Since the incident environmental illumination would be scattered to and from the water bottom, and as we assume a Lambertian bottom surface, this light can be denoted as

$$I_g(\boldsymbol{u}') = T(\theta_r(\boldsymbol{u}'))I_{g0}\,, \tag{10}$$

where I_{g0} is the uniform Lambertian reflection at the bottom, and $T(\theta_r(\boldsymbol{u}'))$, or abbreviated as $T(\boldsymbol{u}')$, is the transmittance into the camera.

It is important to note that this component is negligible when considering reflected observations by ponds and lakes that have sufficient depth. It becomes dominant only for shallow water media such as puddles.

5.4. Scene Radiance and HDR Recovery

To recover the scene radiance from the reflected observation, we first estimate the sum of diffuse bottom surface reflectance I_{g0} and subsurface scattering I_{s0} from each of N pairs of direct and reflected image coordinates $(\boldsymbol{u}_i, \boldsymbol{u}_i')$ and their observations $(I_c(\boldsymbol{u}_i), I_c(\boldsymbol{u}_i'))$ $(i = 1, \ldots, N)$ and average them using Eqs. 4, 9, 10

$$\begin{aligned} I_c(\boldsymbol{u}_i') &= F(\boldsymbol{u}_i')I_c(\boldsymbol{u}_i) + T(\boldsymbol{u}_i')(I_{g0} + I_{s0})\,, \\ I_{g0} + I_{s0} &= \frac{1}{N}\sum_{i=1}^{N}\left\{T(\boldsymbol{u}_i')^{-1}\left(I_c(\boldsymbol{u}_i') - F(\boldsymbol{u}_i')I_c(\boldsymbol{u}_i)\right)\right\}\,. \end{aligned} \tag{11}$$

We can then sift out the scene radiance from the reflected observation of each $\boldsymbol{u}'$ by subtracting these additive components and by undoing Fresnel reflection

$$I_{c0}(\boldsymbol{u}') = F(\boldsymbol{u}')^{-1}\left(I_c(\boldsymbol{u}') - T(\boldsymbol{u}')(I_{g0} + I_{s0})\right) . \quad (12)$$

The recovered scene radiance in the reflected observation is then used for stereo matching with its direct observation.

The reflected observation consists of a darkened scene radiance due to Fresnel reflection combined with scattered and bottom-reflected environmental illumination. The latter components are usually less dominant compared to the Fresnel-reflected scene radiance component. Once we eliminate those components, we are basically left with the scene radiance modulated by the angular-varying Fresnel effect, which reduces the scene radiance more as we look at closer water surface in a general image capture setting. In other words, we are left with the scene radiance captured with a varying neutral density filter, which suggests that we can combine the scene radiance values captured in the direct and reflected observations to estimate high-dynamic range appearance of the scene. Although this is HDR recovery from only two different exposures, as shown in Fig. 3, it lets us recover scene appearance details particularly in saturated regions of the direct view or underexposed regions of the reflected view. In Sec. 7, we demonstrate this HDR scene recovery and show tone-mapped results [3].

5.5. Intrinsic Camera Parameter Estimation

In regular stereo reconstruction, changes in intrinsic camera parameters do not alter the fundamental matrix, which in turns means that the camera intrinsics cannot be recovered. In contrast, in shape from water reflection, the reflected scene radiance is altered by the Fresnel water surface reflection whose magnitude depends on the viewing angle $\theta_r(\boldsymbol{u}')$ (equivalently the incident angle). This suggests that we may estimate intrinsic camera parameters, most notably the focal length f_c, when recovering the reflected scene radiance from the reflected observation.

Eq. 7 indicates that the viewing angle θ_r is also a function of the intrinsic camera matrix A. The focal length f_c together with the environmental illumination components $I_{g0} + I_{s0}$ can be estimated by minimizing errors of Eqs. 11 and 12

$$\underset{f_c, I_{g0}+I_{s0}}{\arg\min} \sum_{\langle \boldsymbol{u}, \boldsymbol{u}' \rangle} \left(E_{c0} + \lambda E_{gt}\right) , \quad (13)$$

$$E_{c0} = \left\| I_{c0}(f_c, \boldsymbol{u}', I_{g0} + I_{s0}) - I_c(\boldsymbol{u}) \right\| , \quad (14)$$

$$E_{gt} = \left\| I_{g0} + I_{s0} - T(\boldsymbol{u}')^{-1}\left(I_c(\boldsymbol{u}') - F(\boldsymbol{u}')I_c(\boldsymbol{u})\right) \right\| , \quad (15)$$

where λ is a weighting parameter. The second term E_{gt} evaluates the uniformity of the estimated environmental illumination.

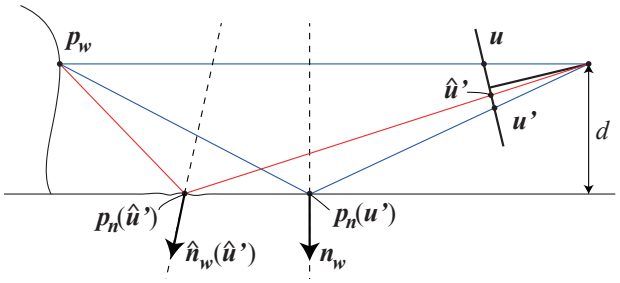

Figure 4. By estimating waves as surface normal variations of the water surface, we can remove the effects of waves and recover the reflection point and image coordinates of the reflected observation for a planar water surface.

6. Wavy Water Reflection

When the water surface has noticeable waves, we simultaneously estimate the geometry of the waves and the scene.

6.1. Wavy Water Surface

As we can safely assume that the wave amplitude is sufficiently small compared to the height of the camera, we can model them as surface normal perturbations to the otherwise flat water surface. This surface normal variation causes changes in projected image coordinates and their radiance (*i.e.*, reflected observations). Using notations depicted in Fig. 4, we model this by expressing the reflected observation for a corresponding pair of direct and reflected observations $(\boldsymbol{u}, \hat{\boldsymbol{u}}')$ using the local normal $\hat{\boldsymbol{n}}_w(\hat{\boldsymbol{u}}')$ of the water surface point from where that reflected observation comes

$$\begin{aligned} I_c(\hat{\boldsymbol{u}}') =& F(\hat{\boldsymbol{u}}', \hat{\boldsymbol{n}}_w(\hat{\boldsymbol{u}}'))I_c(\boldsymbol{u}) \\ &+ T(\hat{\boldsymbol{u}}', \hat{\boldsymbol{n}}_w(\hat{\boldsymbol{u}}'))(I_{g0} + I_{s0}) . \end{aligned} \quad (16)$$

We model the reflected observation as that taken by a collection of reflected viewpoints, *i.e.*, pixel-wise mirrored cameras with mirrored poses $H_w(\hat{\boldsymbol{u}}')$ and translations $\boldsymbol{t}_w(\hat{\boldsymbol{u}}')$

$$\hat{\boldsymbol{u}}' = \lambda_c' A \left(H_w(\hat{\boldsymbol{u}}')\boldsymbol{p}_w + \boldsymbol{t}_w(\hat{\boldsymbol{u}}')\right) , \quad (17)$$

which makes explicit the relationship of the 3D image coordinates $\boldsymbol{u}$ and $\hat{\boldsymbol{u}}'$ in terms of a sum of deformation and disparity on the 2D image plane.

6.2. Waves as 2D Deformations

Our goal is to remove the effects of water surface deformation and associated radiometric change in reflected observation (*i.e.*, image intensity change) by comparing direct and reflected observations independent of the disparity. The main challenge for this is that the angular dependency of the reflected observation only results in a subtle change in image intensity which, in general, is not large enough to robustly separate the water surface normal variation and

Figure 5. Quantitative evaluation using structured-light stereo to acquire ground truth depth. From (a) a single image with bottom surface reflection and waves, we recover (b) HDR appearance and (c,d) dense 3D geometry (shown in two views) which agree well with ground truth as (e) the percentile error map w.r.t. scene depth range shows. Our method also recovers the wave structure as surface normal variations as shown with (f) ten-times amplified normal maps.

disparity. To simultaneously estimate the wavy water surface and the scene geometry, we formulate it as a 2D image alignment and employ prior knowledge about the scene structure and the wavy water surface.

We first project the direct and reflected observations to a common image plane on which we can achieve 2D image alignment between the observations. While we may use arbitrary virtual planes, for simplicity, we use the water surface seen fronto-parallel as the common image plane. Let us denote the common image plane with $\{\boldsymbol{u}'\}$. Note that the 3D image coordinates $\{\boldsymbol{u}'\}$ are unknown because we only know $I_c(\boldsymbol{u})$ and $I_c(\hat{\boldsymbol{u}}')$. We denote the direct observations in the actual input image with $I_D(\boldsymbol{u}')$ and those projected onto the common image plane with $I_c(\boldsymbol{u})$. Similarly, we denote the wave-removed reflected observation (*i.e.*, the estimated reflected scene appearance for a planar water surface) with $I_M(\boldsymbol{u}')$ and its recovered radiometry (*i.e.*, estimated scene radiance) with $I_c(\hat{\boldsymbol{u}}')$.

Our objective is to estimate the waves as local surface normals of the water surface such that

$$I_D(\boldsymbol{u}') = I_M(\boldsymbol{u}')\,. \tag{18}$$

The geometric projection of 3D image coordinates of the direct observation $\boldsymbol{u}$ to that on the common image plane $\boldsymbol{u}'$ in I_D is

$$\boldsymbol{u}' = H_{disp}(\boldsymbol{p}_w)\boldsymbol{u}\,, \tag{19}$$

where $H_{disp}(\boldsymbol{p}_w)$ is a homography matrix determined by the 3D scene point $\boldsymbol{p}_w$ and the local water surface normal $\boldsymbol{n}_w$ for each pixel $\boldsymbol{u}$.

On the other hand, the geometric transformation 3D image coordinates of the reflected observation $\hat{\boldsymbol{u}}'$ to that on the common image plane $\boldsymbol{u}'$ in I_M is

$$\boldsymbol{u}' = g(\hat{\boldsymbol{u}}')\,. \tag{20}$$

We estimate this displacement field g with a generic 2D non-rigid image registration method [30].

6.3. Wavy Water Surface Reconstruction

The estimated displacement field on the common image plane provides correspondences between direct and reflected observations $(\boldsymbol{u}, \boldsymbol{u}')$ and the water surface normals $\boldsymbol{n}_w$ and scene geometry $\boldsymbol{p}_w$ can be recovered from Eq. 19. As depicted in Fig. 4, the displacement estimate tells us that the reflected observation $\hat{\boldsymbol{u}}'$ is moved to $\boldsymbol{u}'$ on the input image for the same scene point $\boldsymbol{p}_w$. The corresponding reflection point on the water surface is moved to $\boldsymbol{p}_n(\boldsymbol{u}')$ after removing the wave

$$\boldsymbol{p}_n(\boldsymbol{u}') = \frac{d}{\cos(\theta_r(\boldsymbol{u}'))}\boldsymbol{v}_c(\boldsymbol{u}')\,. \tag{21}$$

By using the estimated 3D scene coordinates $\boldsymbol{p}_w$, we obtain the surface normal at each reflecting point of the wavy water

$$\hat{\boldsymbol{n}}_w(\hat{\boldsymbol{u}}') = \frac{1}{2}\left(-\frac{\boldsymbol{p}_n(\hat{\boldsymbol{u}}')}{|\boldsymbol{p}_n(\hat{\boldsymbol{u}}')|} + \frac{\boldsymbol{p}_w - \boldsymbol{p}_n(\hat{\boldsymbol{u}}')}{|\boldsymbol{p}_w - \boldsymbol{p}_n(\hat{\boldsymbol{u}}')|}\right). \tag{22}$$

When the waves are erroneously recovered, the recovered scene geometry will subsume the errors in water surface geometry resulting in a wavy 3D scene structure. Since real-world surface geometry is, in general, not wave-like, we can impose a geometric prior on the scene. We employ a piecewise planar geometry prior, that encourages the recovered scene geometry to consist of locally planar surfaces. In particular, we segment the direct observation of the scene into superpixels and impose this piecewise planarity on each superpixel (*i.e.*, a locally connected set of $\{\boldsymbol{u}\}$). We impose this prior in a coarse-to-fine fashion, in which the superpixel size is iteratively refined. This, in effect, allows smoothly curved scene geometry while nudging the wave pattern to be explained by the water surface instead of the scene structure.

Since 2D non-rigid image registration on the common image plane can also erroneously absorb disparity errors as surface normal variations, we also impose a prior on the

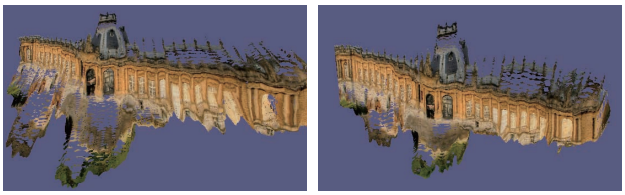

Figure 6. The effect of self-calibration: reconstructed 3D model without (left) and with (right) focal length estimation. The simultaneous focal length estimation (*i.e.*, self-calibration) clearly results in more accurate geometry.

Figure 7. The effect of wave surface recovery: reconstructed 3D model without (left) and with (right) wave geometry estimation. By explicitly recovering the wave surface, our method achieves accurate scene geometry recovery from wavy water reflection.

wave geometry. Inspired by simple computer graphics models for water waves [6, 15], we employ a Fourier domain prior model. The height map $h(\boldsymbol{p}_n, t_0)$ for a 3D wavy water surface point $\boldsymbol{p}_n$ at a single time instance is given as

$$h(\boldsymbol{p}_n) = \boldsymbol{p}_0 + \sum_i a_i e^{i\boldsymbol{k}\boldsymbol{p}_n - \phi_i}\,, \tag{23}$$

where a_i is the amplitude of a Fourier component, and $\boldsymbol{k}$ is a wave vector. The surface normals of the wavy water surface is computed by differentiation of the wave heights. We use a pre-determined number of Fourier components and apply inverse Fourier transform to recover the wave.

7. Experimental Results

We thoroughly evaluate and demonstrate the effectiveness of our method with controlled experiments and with arbitrary images taken in the wild.

To quantitatively evaluate the accuracy of the recovered 3D scene geometry, as shown in the left most of Fig. 5, we setup a dual-camera imaging system in the lab. Although only one camera is used for capturing the input image for our method, we use the other camera and a projector to reconstruct ground truth depth with structured light stereo reconstruction. The two cameras are calibrated beforehand and we also obtained global scale by capturing a checkerboard in the direct and reflected area of the camera image. This provides per-pixel ground truth depth at the main camera. In addition, we create waves by perturbing the water surface to evaluate the method's robustness to reasonably large waves (Fig. 5, the second row).

Fig. 5(c) and (d) show that our method recovers 3D geometry from the single input image (a) with sufficient accuracy as can be seen in the error maps (e) computed against the ground truth. The average depth errors are 6.3% and 3.2% for the calm water surface and the wavy water surface, respectively. Note how dark the reflected observations of the scene in the input images are, which renders any simple appearance adaptation for direct stereo reconstruction impossible. The results show that our method successfully extracts the true scene radiance from the reflected observation which enables robust matching against the direction observation for both cases. The recovered scene appearance, including the saturated intensities of the cup in Fig. 5(b), also show that our method successfully reconstructs high-dynamic range radiance. The missing areas in the reconstruction are occluded from the camera. The simultaneously reconstructed waves shown as an exaggerated surface normal map in Fig. 5(f), also look reasonable, although there are no means to know the ground truth, suggesting successful disentanglement of scene and wave structures.

To quantitatively evaluate the accuracy of self-calibration, *i.e.*, intrinsic parameter recovery, we randomly sampled N pairs of corresponding point pairs $\boldsymbol{u}$ and $\boldsymbol{u}'$ from the input image in the first row of Fig. 5(a), and recovered the focal length using the method described in Sec. 5.5. We found that with more pairs the relative error decreased rapidly and with 100 pairs it already reached less than 5% error. Note that, since our method achieves dense matching in the process of radiometry and geometry reconstruction, many more than 100 point pairs can easily be furnished. In Fig. 6, we also quantitatively demonstrate the effect of self-calibration. The results clearly show that the focal length estimation undoes the skew and results in more visibly accurate geometry.

Fig. 7 shows results of reconstructing the Golden Temple with and without estimation of the waves on the pond reflecting it. The results clearly show that by explicitly modeling the waves as 2D deformations of the proxy plane and imposing natural priors on their shape, they can be disentangled from the target geometry.

Fig. 8 shows comparisons of our method and our implementation of [32]. We compare on two images taken from [32] and two images we have captured. The results clearly show that our method achieves more accurate geometry estimation, in addition to the fundamental difference of also recovering HDR appearance. In general, the recovered geometry by the method in [32] is fragmented, which is a direct result of being inherently restricted to very sparse depth layers due to the heavy reliance on sparse keypoint correspondence pairs that governs the depth range and appearance adaptation. In sharp contrast, by disentangling the radiometric and geometric properties of water reflection, our method is able to achieve dense and clean per-pixel geome-

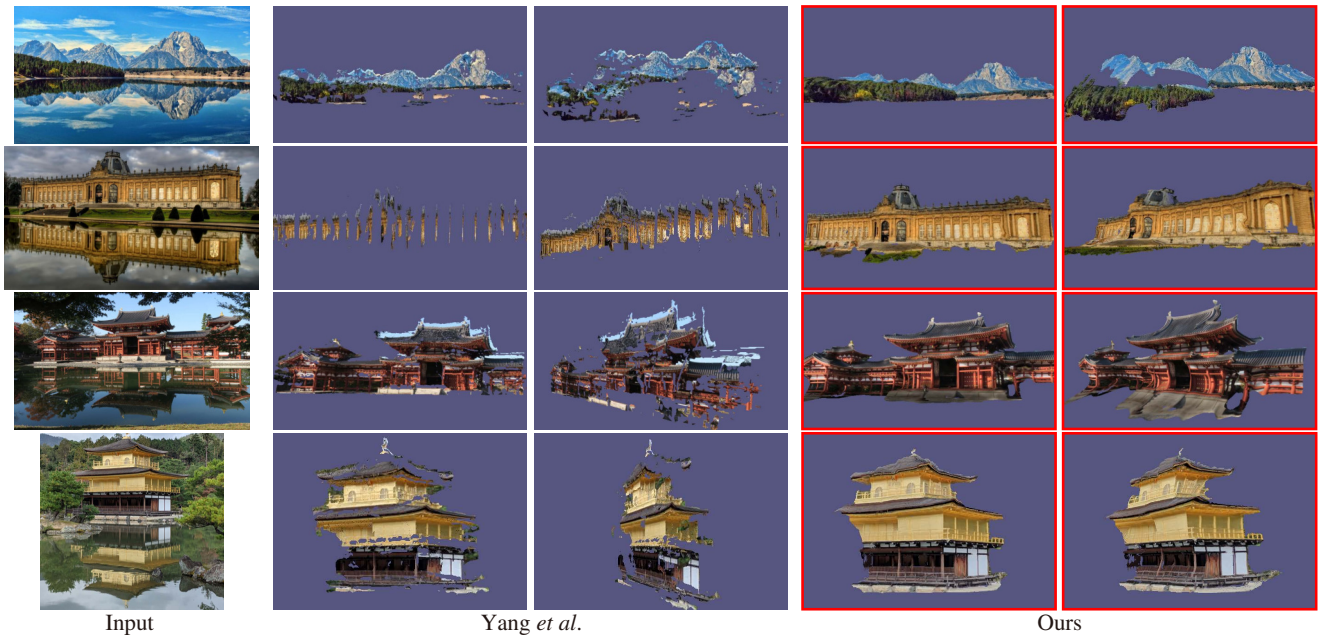

Figure 8. Comparison with Yang *et al*. [32]. From left to right, input images, results by Yang *et al*. [32], and results by our method. Our method successfully reconstructs a dense accurate 3D model of the scene, together with its high-dynamic range appearance. (Image credits for the third and fourth input images: Yuji Wada and Ko Nishino.)

Figure 9. Results on arbitrary images found on the Internet. For each example, we recover an HDR-texture mapped mesh model shown from two different viewing directions. More results can be found in the supplementary video. The results demonstrate the effectiveness of our method on arbitrary already-taken images. (Image credits for the left and right input images: Andrés Nieto Porras and Edwin Giesbers.)

try reconstruction.

We apply our method to various images either taken by our phone cameras or found on the Internet. Fig. 9 shows the single input images and recovered 3D models with high-dynamic range appearance. The supplementary video contains more results. The results show that the 3D scene structure can be recovered despite waves and complex water surface reflection. It is also interesting to see how the method applies to a wide range of scene scales, ranging from a small bird to a large architecture. Images taken with long focal length tend to result in flatter surface with limited depth variation as one expects. The results also include various types of water reflection, ranging from a puddle to a lake demonstrating its successful application to images truly taken in the wild.

8. Conclusion

In this paper, we introduced appearance and shape from water reflection to recover 3D geometry and high-dynamic range appearance of a scene from an image capturing both direct and water-reflected views in a single exposure. The method can recover camera parameters and waves in addition to the scene structure and appearance, enabling its application to unconstrained, already-taken images. Experimental results demonstrate its robustness to waves and its effectiveness when applied to arbitrary images taken in the wild. We believe the method has strong implications in a wide range of domains, not just in vision and graphics, but also in photography as a new visual media, as well as in image forensics analysis in which direct and water-reflected geometry and radiometry, even with waves, can now be used as visual cues of image tampering.

Acknowledgement This work was in part supported by JSPS KAKENHI 15H05918, 17K20143, 18K19815, and 26240023.

References

[1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *TPAMI*, 14(2):99–106, 1992. 2

[2] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *IJCV*, 35(2):175–196, 1999. 2

[3] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced High Dynamic Range Imaging (2nd Edition)*. AK Peters (CRC Press), Natick, MA, USA, July 2017. 5

[4] M. Baradad, V. Ye, A. B. Yedidia, F. Durand, W. T. Freeman, G. W. Wornell, and A. Torralba. Inferring light fields from shadows. In *Proc. CVPR*, pages 6267–6275, 2018. 2

[5] K. L. Bouman, V. Ye, A. B. Yedidia, F. Durand, G. W. Wornell, A. Torralba, and W. T. Freeman. Turning corners into cameras: Principles and methods. In *Proc. ICCV*, 2017. 2

[6] E. Darles, B. Crespin, D. Ghazanfarpour, and J. Gonzato. A survey of ocean simulation and rendering techniques in computer graphics. *Computer Graphics Forum*, 30(1):43–60, 2011. 7

[7] K. Forbes, F. Nicolls, G. D. Jager, and A. Voigt. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *Proc. ECCV*, 2006. 2

[8] J. Gluckman and S. K. Nayar. Catadioptric stereo using planar mirrors. *IJCV*, 44(1):65–79, 2001. 2

[9] H. Ha, S. Im, J. Park, H. Jeon, and I. S. Kweon. High-quality depth from uncalibrated small motion clip. In *Proc. CVPR*, pages 5413–5421, 2016. 2

[10] S. Im, G. Choe, H. Jeon, and I. S. Kweon. Depth from accidental motion using geometry prior. In *Proc. ICIP*, pages 4160–4164, 2015. 2

[11] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proc. ACM SIGGRAPH*, pages 511–518, 2001. 4

[12] N. Kong, Y. Tai, and J. S. Shin. A physically-based approach to reflection separation: From physical modeling to constrained optimization. *TPAMI*, 36(2):209–221, Feb 2014. 2

[13] D. Lanman, D. Crispell, and G. Taubin. Surround structured lightning: 3-d scanning with orthographic illumination. In *CVIU*, pages 1107–1117, November 2009. 2

[14] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. In *eccv*, pages 602–613, 2004. 2

[15] H. Li, H. Yang, C. Xu, and J. Zhao. Fast global illumination of dynamic water surface based on two stage rendering. *Cluster Computing*, Mar 2018. 7

[16] Y. Mukaigawa, S. Tagawa, J. Kim, R. Raskar, Y. Matsushita, and Y. Yagi. Hemispherical confocal imaging using turtleback reflector. In *Proc. ACCV*, 2011. 2

[17] S. A. Nene and S. K. Nayar. Stereo with mirrors. In *Proc. ICCV*, pages 1087–1094, 1998. 2

[18] K. Nishino and S. Nayar. The World in Eyes. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 444–451, 2004. 2

[19] K. Nishino and S. K. Nayar. Corneal imaging system: Environment from eyes. *IJCV*, 70(1):23–40, 2006. 2

[20] I. Reshetouski, A. Manakov, H.-P. Seidel, and I. Ihrke. Three-dimensional kaleidoscopic imaging. In *Proc. CVPR*, pages 353–360, 2011. 2

[21] B. Sarel and M. Irani. Separating transparent layers through layer information exchange. In *eccv*, pages 328–341, 2004. 2

[22] D. Scaramuzza, A. Martinelli, and R. Siegwart. A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *Fourth IEEE International Conference on Computer Vision Systems*, pages 45–45, 2006. 2

[23] S. Shwartz, Y. Y. Schechner, and M. Zibulevsky. Blind separation of convolutive image mixtures. *Neurocomputing*, 71(10):2164 – 2179, 2008. 2

[24] G. Somanath, M. V. Rohith, and C. Kambhamettu. Single camera stereo system using prism and mirrors. In *Advances in Visual Computing*, pages 170–181, 2010. 2

[25] P. Sturm and J. P. Barreto. General imaging geometry for central catadioptric cameras. In *Proc. ECCV*, pages 609–622, 2008. 2

[26] T. Tahara, R. Kawahara, S. Nobuhara, and T. Matsuyama. Interference-free epipole-centered structured light pattern for mirror-based multi-view active stereo. In *Proc. 3DV*, pages 153–161, 2015. 2

[27] K. Takahashi, A. Miyata, S. Nobuhara, and T. Matsuyama. A linear extrinsic calibration of kaleidoscopic imaging system from single 3d point. In *Proc. CVPR*, 2017. 2

[28] A. Torralba and W. T. Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. In *Proc. CVPR*, pages 374–381, 2012. 2

[29] Y. Tsin, S. B. Kang, and R. Szeliski. Stereo matching with reflections and translucency. In *Proc. CVPR*, volume 1, 2003. 2

[30] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supplement 1):S61 – S72, 2009. Mathematics in Brain Imaging. 6

[31] W. Xiong, H. S. Chung, and J. Jia. Fractional stereo matching using expectation-maximization. *TPAMI*, 31(3):428–443, March 2009. 2

[32] L. Yang, J. Liu, and X. Tang. Depth from water reflection. *IEEE Trans. on Image Processing*, 24(4):1235–1243, 2015. 1, 2, 7, 8

[33] F. Yu and D. Gallup. 3d reconstruction from accidental motion. In *Proc. CVPR*, 2014. 2