A DIFFERENTIAL-GEOMETRIC APPROACH FOR GLOBALLY SOLVING A NON-CONVEX, DISCONTINUOUS DEPTH ESTIMATION PROBLEM FOR PLENOPTIC CAMERA IMAGES

Isaac J. Sledge[1], *Member, IEEE* José C. Príncipe[2,3], *Life Fellow, IEEE*

[1] Advanced Signal Processing and Automated Target Recognition Branch, Naval Surface Warfare Center—Panama City
[2] Department of Electrical and Computer Engineering, University of Florida
[3] Department of Biomedical Engineering, University of Florida

ABSTRACT

In this paper, we address the problem of estimating a scene's three-dimensional geometry from plenoptic camera images. Existing approaches for this problem have emphasized the development of sharpness and contrast measures for distinguishing between in-/out-of-focus image regions. The ways in which these measures are aggregated can yield erroneous, localized distance fluctuations, though. To deal with such fluctuations, post-processing smoothing techniques can be applied. However, they may remove fine-scale, non-erroneous depth structures and edges. Here, we propose a non-convex, discontinuous cost-function that simultaneously combines and regularizes sharpness and contrast so that valid depth transitions are better preserved. We implicitly convert this function into one that is continuous and (quasi-)convex by optimizing it on the non-positively-curved Riemannian manifold of depth maps with a learned metric.

Index Terms—Depth from focus, depth estimation, manifold optimization, image processing

1. INTRODUCTION

Inferring the three-dimensional shape of objects from two-dimensional information is a fundamental task for many problems. Out of the myriad shape recovery approaches, those that rely on focus analysis have received substantial attention. These approaches can be divided into two categories: depth from focus [1, 2] and depth from defocus [3–5]; we focus on the former category in this paper.

In depth-from-focus, distance information is estimated from image sequences of the same scene captured with different degrees of focus. Approaches to the problem can exploit properties of camera models to analytically resolve scene distances. In practice, though, investigators may not have access to the necessary camera parameters; issues with measurement noise may also be encountered, which can affect the depth estimates. It is thus common to instead rely on a two-step, approximate procedure to overcome such issues. The first step entails discerning cues about depth from local focus variation, which detects the in-focus regions of each image in a focal stack. Such variations can be found using one of many available sharpness/focus measures [6–10] that respond to regions with stark changes in texture; we refer to [11] for comparisons of these measures and their performances. The second step involves aggregating the in-focus information from the entire focal stack to estimate the depth of every point in the scene. This reconstruction phase often requires the model-based interpolation of the focal measure [12]. Most of these models are heuristic in nature.

Depth-from-focus estimates may contain spurious artifacts. The estimates can sometimes be improved by either post-processing the focus measure [13] or by exploiting defocus information [14–16]. As an example, Muhammad et al. [17] advocate discarding portions of the scene with high depth variations. Their reasoning is that such regions arise due to an inaccurate computation of the focus measure. The discarded regions are then recovered via spatial interpolation. A somewhat similar approach is taken by Pertuz, Puig, and Garcia in [18]. In [19], Gaganov and Ignatenko utilize a spatial, random-field-based model to smooth depth maps in low-reliability areas. Both types of approaches assume that details from highly-textured areas can be used to better infer depth map. However, with the exception of [18], it is not clear how to distinguish between low- and high-reliability areas, especially when one or more sharpness/focus measures are performing poorly. Substantial errors can therefore be unnecessarily propagated into the produced depth maps.

Another common technique to enhance the quality of the focus volume, after the depth-estimation process, is to combine sharpness or contrast values within small-scale neighborhoods [20, 21]. This fusion of this information serves to smooth some of the erroneous depth fluctuations. Unfortunately, the efficacy of this technique is greatly affected by the neighborhood size. Malik and Choi [20] determined that large neighborhoods tend to remove fine-scale depth structures and valid, sudden transitions between regions of differing depths. They hence suggested relying on small neighborhoods. While such neighbors preserve such structures rather well, they often do not adequately suppress depth noise.

We posit that modifications to the depth maps should occur during the aggregation phase, not after it, so as to reduce the potential for perturbing the depths in high-reliability regions. Toward this end, we introduce a framework for the depth-from-focus problem for plenoptic camera images. This framework relies on a two-term cost function. The first term, a data fidelity component, quantifies depth using the sum-modified Laplacian [6] sharpness measure. The second term, a regularization component, penalizes depth estimates that fluctuate greatly over local neighborhoods. It utilizes the discrete, isotropic total variation for smoothing depth while preserving edge information.

The variational objective function that we define is innately discontinuous and non-convex. Discontinuity of the functional is due to the regularization component, while non-convexity can stem from the data fidelity component. To deal with these issues, we optimize the cost over Riemannian manifolds of matrices with non-positive curvature. Depending on the chosen metric, the non-convex cost function can become either quasi-convex or convex. Additionally, on such manifolds, we can separate the non-smooth parts of the functional into differentiable and non-differentiable components. We can then define fixed-point iterations of proximity operators to handle the non-differentiable components. Due to the convexity of both components, we are assured that the global-best depth map, as characterized by our cost, will be uncovered. The theoretical treatment of this process is the main contribution of our paper.

2. METHODOLOGY

In what follows, $\mathcal{M}$ is a connected, embedded manifold that is a smooth subset of a vector space included in the set of matrices $\mathbb{R}^{m\times n}$. The tangent space for $\mathcal{M}$ is denoted by $T_{p_i}\mathcal{M}$ for a point $p_i \in \mathcal{M}$. The associated co-tangent space is denoted by $T^*_{p_i}\mathcal{M}$.

On each tangent space, we can define a family of inner products that smoothly varies: that is, $p_i \mapsto \langle x(p_i), y(p_i)\rangle_{p_i}$ is a smooth function for vector fields x, y. Such families give rise to a Riemannian metric. When $\mathcal{M}$ is endowed with a Riemannian metric g, it is referred to as a Riemannian manifold.

For $x_i \in \mathcal{M}$, we have a curvature operator $\kappa_{v_i} : T_{x_i}\mathcal{M} \to T_{x_i}\mathcal{M}$ for an non-zero vector $v_i \in T_{x_i}\mathcal{M}$; we assume that $\mathcal{M}$ is additionally complete. From the Bianchi identities, κ_{v_i} is a linear operator. It is negative semi-definite for each v_i if the sectional curvature of $\mathcal{M}$ is non-positive. In such a case, we are dealing with a manifold with non-positive curvature.

This work was funded by ONR grants N00014-15-1-2013 and N00014-14-1-0542. The first author was additionally funded by a UF Research Fellowship, a UF Robert C. Pittman Fellowship, and an ASEE NREIP Fellowship.

2.1 Cost Function

We propose to address the depth-from-focus problem by minimizing a joint cost functional $f:\mathcal{M}\to\mathbb{R}$ that converts a depth map defined on a non-positively-curved, Riemannian matrix manifold to cost, or energy, magnitudes. This cost is a combination of two components: one for data fidelity f_1 and another for depth regularization f_2. That is, $f(x_i)=f_1(x_i)+\alpha f_2(x_i)$, $x_i\in\mathcal{M}$, for some regularization weight coefficient $\alpha\geq 0$.

The data fidelity term will be used to quantify the goodness of the depth map according to some chosen sharpness, or contrast measure, applied to each image in the focal stack obtained from a plenoptic camera. A thorough comparison of existing contrast measures [11] indicates that the sum-modified Laplacian [6] performs the best. We opted to use this contrast measure as our data fidelity term

$$f_1(x_i) = -\textstyle\sum_{j,k,r} c_{(j,k,r)}x_{i,(j,k)},\quad c_{(j,k,r)}=[D^2y_r]_{(j,k)}$$

where $D^2y_r:\mathcal{M}\to\mathcal{N}$ is a forward, second-order finite-difference operator that maps to the manifold of $\mathcal{N}\simeq(\mathbb{R}^2)^{m\times n}$ vector fields that is diffeomorphic to the reals. The element $y_{(r)}\in\mathcal{M}$ represents the rth focal stack image.

The regularization component enforces a smoothness constraint on the scene reconstruction. This component should incorporate any prior knowledge about the expected properties of the associated depth map. Here, we utilize the discrete, isotropic total variation regularizer

$$f_2(x_i) = \max_{u\in\mathcal{N},\mathcal{N}\simeq(\mathbb{R}^2)^{m\times n}} \{\langle Dx_i,u\rangle : -1\leq u_{(j,k)}\leq 1,\ \forall j,k\}$$

where $Dx_i:\mathcal{M}\to\mathcal{N}$ is a forward, first-order finite-difference operator. Our motivation for utilizing total variation is due to its well-known edge preservation properties: we want to ensure that widespread jumps in depth are not overly smoothed while suppressing noisy, localized fluctuations in distance.

2.2 Cost Optimization and Analysis

By default, the data-fidelity component may not be continuous. To deal with this issue, we can consider multi-order polynomial approximations. In utilizing such approximations, though, we potentially forgo convexity. On the other hand, while the regularization component is convex, it is not inherently smooth.

In the optimization literature, there are schemes that can address functionals with components of varying smoothness and convexity. A prominent example involves using forward-backward splitting [22]. Forward-backward splitting entails performing a forward gradient step on the smooth, convex functional components. A backward gradient step involving sub-gradients is then carried out on the non-smooth, potentially non-convex components. Such schemes are provably convergent to critical points, commonly under mild conditions.

There are some drawbacks to such schemes. A typical means of dealing with non-convexity in forward-backward splitting is to linearize the associated components. However, linearization can drastically alter the associated solutions of the cost functional. In our problem, linearization of the regularization component can yield depth maps that contain a series of sharp, unnecessary fluctuations in depth. We are trying to avoid this exact situation: we would like there to be gradual transitions in depth over neighboring regions with similar depth estimates. Sharp transitions should only occur in regions where there is pronounced, widespread change in depth.

To deal naturally with non-convexity and non-smoothness, we will optimize our functional on non-positively-curved Riemannian manifolds endowed with a learned metric. We generate candidate depth maps $x_i^k\in\mathcal{M}$, $k=1,2,\ldots$ according to the following proximal-gradient iteration

$$x_i^{k+1}\leftarrow \operatorname*{arg\,inf}_{x_j\in\mathcal{M}} f(x_j)+\lambda^k\langle \exp_{x_j}^{-1}(x_i^k),\exp_{x_j}^{-1}(x_i^k)\rangle^2/2$$

for the inverse exponential map. Here, the step-length variables $\lambda^k>0$, $\lim_{k\to\infty}\lambda^k\to 0$, are each positive numbers for all time steps k. Large values of λ^k yield more changes in the intermediate depth maps than small values. Hence, the sequence should move more quickly toward a stationary point for larger magnitudes. However, after a certain threshold is reached for λ^k, the iteration destabilizes and may not reach the global optimum. We therefore implicitly assume that λ^k is bounded from above so as to prevent this occurrence.

In what follows, we demonstrate that this proximal-gradient iteration converges to global minimizers of the cost function on non-positively-curved, complete Riemannian manifolds. Proofs of these claims are given in an online appendix[1].

2.2.1 Convexity for Non-Positively-Curved Manifolds

In this sub-section, we draw some connections between Euclidean spaces and complete, Riemannian manifolds with non-positive sectional curvature. By establishing these connections, we can specify the notion of convex functions on such manifolds.

Proposition 2.1 [23]: Let $x_i\in\mathcal{M}$, where $\mathcal{M}$ has non-positive sectional curvature and is Riemannian. Let $\exp_{x_i}:T_{x_i}\mathcal{M}\to\mathcal{M}$ be the exponential map. Then:

(i) $\|d\exp_{x_i}(\beta)\|\geq\|\beta\|$, $\beta\in T_{v_i}T_{x_i}\mathcal{M}$, where $v_i\in\mathcal{T}_{x_i}\mathcal{M}$. We assume that $T_{v_i}T_{x_i}\mathcal{M}$ inherits the Euclidean metric from the tangent space $T_{x_i}\mathcal{M}$.

(ii) If $\gamma:\mathbb{R}\to T_{x_i}\mathcal{M}$ is a smooth curve, then its length is bounded above by the length of $\exp_{x_i}\circ\gamma$. If $\mathcal{M}$ is simply connected, then $d(\exp_{x_i}(v_i),\exp_{x_i}(v_i'))$ is bounded below by the distance between any $v_i,v_i'\in T_{x_i}\mathcal{M}$.

Proposition 2.2 [23]: Let $x_i\in\mathcal{M}$, where $\mathcal{M}$ has non-positive sectional curvature. Then:

(i) $\exp_{x_i}:T_{x_i}\mathcal{M}\to\mathcal{M}$ is a covering map.

(ii) If $\mathcal{M}$ is simply connected, then $\exp_{x_i}$ is a diffeomorphism. More specifically, if $p_i,q_i\in\mathcal{M}$ are distinct, then there is a unit-speed geodesic $\gamma_{p_i,q_i}:\mathbb{R}\to\mathcal{M}$ with $\gamma_{p_i,q_i}(0)=p_i$ and $\gamma_{p_i,q_i}(d(p_i,q_i))=q_i$. This geodesic is unique.

Propositions 2.1 and 2.2 can be used to show that a Riemannian manifold $\mathcal{M}$ with non-positive curvature is diffeomorphic to $\mathbb{R}^{m\times n}$. Thus, $\mathcal{M}$ has the same topology as $\mathbb{R}^{m\times n}$. These manifolds hence share some similar geometric properties, which allows for the construction of convex sets on them.

The notion of convex sets on manifolds of non-positive curvature is somewhat analogous to that of convex sets for Euclidean spaces. The subset $\mathcal{U}$ of $\mathcal{M}$ is said to be convex if a geodesic segment with end-points in $\mathcal{U}$ is itself entirely contained in $\mathcal{U}$. A function $f:\mathcal{M}\to\mathbb{R}$ is said to be convex if, for any geodesic segment on an open convex set $\mathcal{U}$, the composition $f\circ\gamma:\mathbb{R}\to\mathbb{R}$ is convex.

The following result characterizes convexity for differentiable functions defined on non-positively-curved manifolds. A consequence of this proposition is that the critical points for any convex function are global optima.

Proposition 2.3 [23]: Let $\mathcal{U}$ be an open convex subset of a non-positively-curved, Riemannian manifold $\mathcal{M}$. Let $f:\mathcal{M}\to\mathbb{R}$ be a differentiable function on $\mathcal{U}$. The function f is convex on $\mathcal{U}$ for $p_i,q_i\in\mathcal{U}$ if and only if $f(q_i)-f(p_i)\geq\langle\mathrm{grad}f(p_i),\exp_{q_i}^{-1}(q_i)\rangle$.

We can alternatively use the monotonicity of a function's gradient to obtain that it is convex on non-positively-curved manifolds. Such a result will be particularly important for simplifying the convergence analysis of our framework: we define an iteration that is the sum of a cost functional and a strongly monotone vector field.

Let $\mathcal{U}$ be an open, convex set on the manifold $\mathcal{M}$. Given a vector field y defined on $\mathcal{M}$, this vector field is said to be monotone on $\mathcal{U}$ if

$$\langle\exp_{q_i}^{-1}(p_i),\mathrm{para}_{q_i,p_i}^{-1}y(p_i)-y(q_i)\rangle\geq 0,$$

for all $p_i,q_i\in\mathcal{U}$. Here, $\mathrm{para}_{q_i,p_i}^{-1}$ is the inverse parallel transport along the geodesic joining points q_i to p_i. A vector field y is strongly monotone [?] if, for $\alpha>0$,

$$\langle\exp_{q_i}^{-1}(p_i),\mathrm{para}_{q_i,p_i}^{-1}y(p_i)-y(q_i)\rangle\geq\alpha d^2(p_i,q_i),$$

[1]Temporary web link for the online appendix: https://www.dropbox.com/s/ohpit2rh9safjrz/Sledge-ICASSP-2019-2col-appendix.pdf?dl=0 (A copy of the paper and the appendix will be uploaded to arXiv upon acceptance.)

for any $p_i, q_i \in \mathcal{U}$.

Proposition 2.4: Assume that $\mathcal{U} \subset \mathcal{M}$ is an open, convex set. For any differentiable function $f : \mathcal{M} \to \mathbb{R}$ on a non-positively-curved, Riemannian manifold, we have that:

(i) The function f is convex on the set $\mathcal{U}$ if and only if $\mathrm{grad}(f)$ is monotone on $\mathcal{U}$.

(ii) The function f is strongly convex on $\mathcal{U}$ if and only if $\mathrm{grad}(f)$ is strongly monotone on $\mathcal{U}$.

2.2.2 Derivatives for Non-Positively-Curved Manifolds

In this sub-section, we recall the notion of the directional derivative of a manifold-based convex function and discuss some of its properties [24]. We also recall the notion of sub-gradients and sub-differentials for a certain class of convex functions.

Assume that we have an open, convex subset $\mathcal{U}$ of a non-positively-curved $\mathcal{M}$ and a convex function $f : \mathcal{M} \to \mathbb{R}$ on it. Let γ be a geodesic on $\mathcal{U}$ such that $\gamma(0) = p_i$ and $\dot{\gamma}(0) = q_i$ for $p_i \in \mathcal{U}$ and $q_i \in T_{p_i}\mathcal{M}$. The directional derivative [25] of f at p_i in the direction of q_i is given by $f'(p_i, q_i) = \inf_{t>0} f(\gamma(t))/t - f(p_i)/t$. From the convexity of $f \circ \gamma$, the directional derivative is non-decreasing.

Let $f : \mathcal{M} \to \mathbb{R}$ be a convex function. A one-form $w_{p_i} \in T^*_{p_i}\mathcal{M}$ is called the sub-gradient of f at p_i if $f(q_i) \geq f(p_i) + w_{p_i}(\dot{\gamma}_{p_i,q_i}(0))$ for all $p_i, q_i \in \mathcal{M}$. Here, γ_{p_i,q_i} is the geodesic such that

$$\gamma_{p_i,q_i}(0) = p_i \text{ and } \gamma_{p_i,q_i}(1) = q_i, \text{ with } \dot{\gamma}_{p_i,q_i}(0) \in T_{p_i}\mathcal{M}.$$

The set of all sub-gradients of f at p_i is called the sub-differential of f at p_i and is denoted by $\mathrm{sub}(f(p_i))$. The multi-form map $\mathrm{sub}(f) : p_i \mapsto \mathrm{sub}(f(p_i))$ is referred to as the sub-differential of f.

Proposition 2.5: Let $\mathcal{U}$ be an open, convex subset of $\mathcal{M}$. Let $f_i : \mathcal{M} \to \mathbb{R}$ be a differentiable, convex function on $\mathcal{U}$ for $i = 1, 2, \ldots$ If $f : \mathcal{M} \to \mathbb{R}$ is given by $f(x_j) = \max_i f_i(x_j)$, then the sub-differential of $f(x_j)$ is given by $\mathrm{conv}(\mathrm{grad}(f_i))$, or, rather the set of all $q_i \in T_{p_j}\mathcal{M}$ such that

$$q_j = \textstyle\sum_i a_i \mathrm{grad}(f_i(p_j)), \;\; \sum_i a_i = 1,\; a_i \geq 0$$

where the index i is such that $f(p_j) = f_i(p_j)$. The variable p_j attains a minimum of the function f if and only if

$$\textstyle\sum_i a_i \mathrm{grad}(f_i(p_j)) = 0, \;\; \sum_i a_i = 1.$$

For our analysis, we will also need a version of the directional derivative for a locally Lipschitz function that is not necessarily convex. Let $\mathcal{U}$ be an open, convex subset of $\mathcal{M}$. On this domain, we define a locally Lipschitz function $f : \mathcal{M} \to \mathbb{R}$. The generalized directional derivative of f at $p_i \in \mathcal{U}$ and in the direction $v_i \in T_{p_i}\mathcal{M}$ is given by

$$f'(p_i, q_i) = \lim_{t \to 0} \sup_{v_i \to p_i} f(\exp_{q_i}(t(D\exp_{p_i})_{\exp^{-1}_{p_i}(q_i)}(v_i)))/t - f(q_i)/t.$$

We refer to the differential of $\exp_{p_i}$ at $\exp^{-1}_{p_i}(q_i)$ by $(D\exp_{p_i})_{\exp^{-1}_{p_i}}$.

Associated with a locally Lipschitz function is a generalization of the notion of the sub-differential. Let $\mathcal{U}$ be an open, convex subset of a non-positively-curved $\mathcal{M}$. On this domain, we define a locally Lipschitz function $f : \mathcal{M} \to \mathbb{R}$. The generalized sub-differential $\mathrm{gsub}(f(p_i))$ is defined by the set of all $v_i \in T_{p_i}\mathcal{M}$ such that $\langle v_i, w_i \rangle$ is less than or equal to $f'(p_i, w_i)$ for all $w_i \in T_{p_i}\mathcal{M}$, $p_i \in \mathcal{M}$. Here, $f'(p_i, w_i)$ is the generalized directional derivative.

Proposition 2.6: Let $\mathcal{U}$ be an open, convex subset of a non-positively-curved $\mathcal{M}$. Let $f_i : \mathcal{M} \to \mathbb{R}$ be a continuously differentiable function on $\mathcal{U}$. If $f : \mathcal{M} \to \mathbb{R}$ is given by $f(x_j) = \max_i f_i(x_j)$, then f is locally Lipschitz on $\mathcal{U}$. For each $p_j \in \mathcal{U}$, $\mathrm{conv}(\mathrm{grad}(f_i(p_j)))$ is a subset of the generalized sub-differential $\mathrm{gsub}(f(p_j))$. The index i is such that $f(p_j) = f_i(p_j)$.

2.2.3 Depth-Map Global Convergence Analysis

In this sub-section, we verify that the depth-map iteration will eventually converge to the global-best depth map even if our cost functional is originally non-convex before the choice of a suitable Riemannian metric. Our first objective is to demonstrate the convexity of the iteration process. Convexity of the iteration relies on using a special class of functionals to recreate our non-convex cost function.

Proposition 2.7: Let $\mathcal{U}$ For any $x_i^k \in \mathcal{M}$, the cost function $f(x_i^k) + \lambda^k \langle \cdot, \exp^{-1}_{x_i}(x_i^k) \rangle^2/2$, for $\lambda^k > 0$, is strongly convex in the open, convex subset $\mathcal{U}$ of $\mathcal{M}$; here $\mathcal{M}$ is Riemannian and has non-positive curvature.

The proof of this claim relies on showing that $\mathrm{grad}(f_i(x_i^k))$ is locally Lipschitz on $\mathcal{U}$, according to proposition 2.6. Hence, the cost function is strongly monotone, which implies convexity via proposition 2.4.

Next, we need that the sequence of intermediate depth maps $x^k, x^{k+1}, \ldots$ is well defined. That is, $x^{k+1} \in \mathcal{M}$, $\forall k$, exists and is unique for manifolds with non-positive curvature.

Proposition 2.8: For any depth map initialization $x_i^0 \in \mathcal{M}$, the iteration

$$x_i^{k+1} \leftarrow \arg\inf_{x_j \in \mathcal{M}} f(x_j) + \lambda^k \langle \exp^{-1}_{x_j}(x_i^k), \exp^{-1}_{x_j}(x_i^k) \rangle^2/2$$

is well defined.

Proposition 2.8 is then used, along with propositions 2.6 and 2.7, to prove that the sequence of depth maps yields monotonically non-increasing costs.

Proposition 2.9: Assume that $\{x_i^k\}_k$ is a sequence of intermediate depth maps generated by the above iteration scheme. As well, assume that the positive variable sequence $\{\lambda^k\}_k$ is strictly bounded below by the Lipschitz constant of the cost functional. Let $\mathcal{S}_{p_i,q_i}(f(q_i))$ the set of all $p_i \in \mathcal{M}$ such that

$$f(p_i) \leq f(q_i), \quad \inf_{p_i \in \mathcal{M}} f(p_i) < f(q_i).$$

If for all $p_i \in \mathcal{S}_{p_i,q_i}(f(q_i)) \backslash \mathcal{S}_{p_i,q_i}(c)$ and $y(p_i) \in \mathrm{gsub}(f(p_i))$ we have that $\langle y(p_i), y(p_i) \rangle > 0$, then the sequence of intermediate depth maps belongs to $\mathcal{S}_{p,q}(c) \subset \mathcal{U}$.

Lastly, propositions 2.5, 2.7, 2.8, and 2.9 are employed to prove convergence of the sequence.

Proposition 2.10: Let $\mathcal{U}$ be an open convex subset of $\mathcal{M}$. Let $f_j : \mathcal{M} \to \mathbb{R}$ be a differentiable function on $\mathcal{U}$ that is continuous on the closure of $\mathcal{U}$. We define $f(x_i) = \max_j f_j(x_i)$, for which we assume that $\inf_{x_i \in \mathcal{M}} f(x_i) > -\infty$. For any initial depth map, the sequence generated by the above iteration resides in $\mathcal{S}_{p_i,q_i}(f(q_i)) \subset \mathcal{U}$. Moreover, one of the following is true:

(i) The sequence of depth maps $\{x_i^k\}_k$ is finite and x_i^k is a stationary point of the objective functional f.

(ii) The sequence of depth maps $\{x_i^k\}_k$ is infinite and any accumulation point of the sequence is a stationary point of the objective functional f.

The above analysis relies on the fact that the cost function is convex for the appropriate choice of Riemannian metric. Such a metric can be learned in a data-driven fashion on the manifold of Riemannian metrics $\mathrm{met}(\mathcal{M})$ endowed with the Lebesgue metric. In particular, for an initial guess of the metric $g_0 \in \mathrm{met}(\mathcal{M})$, we find a correction $h^k \in T_{g^k}\mathrm{met}(\mathcal{M})$, $k = 0, 1, \ldots$, where $\nabla_{\dot{\gamma}} \langle \nabla f, \dot{\gamma} \rangle_{g^k} \geq 0$ and $\langle \mathrm{grad}\, f(x_i; g^k), h^k \rangle < 0$. We then update this metric, in an alternating manner with the proximal-gradient iterations, via

$$g^{k+1} \leftarrow \exp_{g^k}(\theta^{i^k} h^k),$$
$$i^k \geq 0 \text{ such that } f(x_i; \exp_{g^k}(\theta^{i^k})) - f(x_i; g^k) \leq \beta\theta^{i^k}$$

for $0 \leq \beta, \theta \leq 1$.

3. EXPERIMENTS

In the previous section, we developed a framework for finding globally-optimal depth-from-focus solutions. Given a focal stack, this framework relies on a two-part cost function to recover depth estimates and remove abrupt, localized transitions in them.

We now assess our depth from focus framework using focal stacks taken from a Lytro plenoptic camera. Results for four different scenes are presented in figures 1(a)–1(d). The left-hand sides of these figures highlight instances of the focal stacks. The number of images in the focal stacks ranged from eight to fifteen, depending on the scene, and

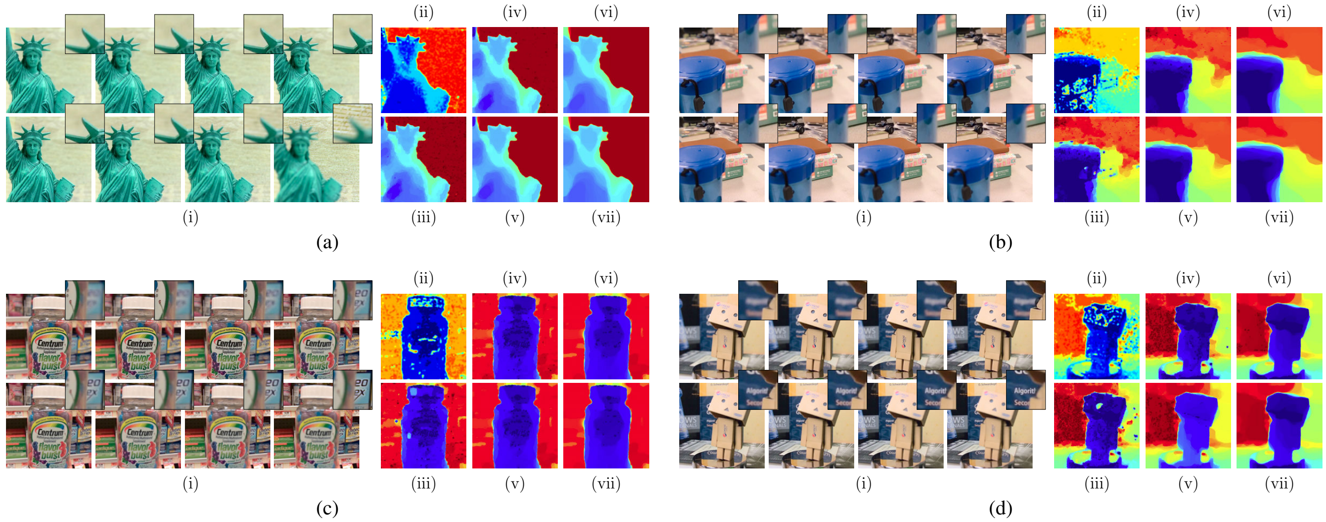

Figure 1: Depth-from-focus results for four scenes. (i) Examples of images from the focal stack. The sub-images in the right-hand corners highlight a particular portion of the image as the focal length is changed. The images are ordered such that the objects nearest to the camera are in focus for the top-left image. The objects furthest away from the camera are in focus for the bottom-right image. Objects that are intermediate distances away from the camera become focused as the focal plane sweeps through these two extremes. (ii) The depth from focus result for the sum-modified Laplacian measure. (iii) The depth from focus results for our variational framework when not learning a metric, which leads to a non-convex problem. (iv)–(vii) Results when learning a metric for different regularization weights. Raising the regularization weight α increasingly removes localized depth fluctuations. The regularization weights are: (iv) 0.1, (v) 0.3, (vi) 0.6, and (vii) 0.9. The depth-image color scheme is such that dark blue corresponds to distances closest to the camera and dark red to distances that are furthest away.

their resolution was 540×540. On the right-hand sides of these figures are the returned depth maps from the classical sum-modified Laplacian method. We also show results for our framework when a metric is not learned, leading to a non-convex cost function, and when one is, yielding a convex cost. For the latter case, we varied the regularization weight α to gauge its effect on the distance smoothing.

In figures 1(a)–1(d), it is apparent that the basic sum-modified Laplacian framework succeeds in finding the general depth structure. In figure 1(a)(ii), for instance, it captures that the chest, arms, and face of the Statue of Liberty figurine are closer than the neck and part of the tablet. However, this framework often introduces distance fluctuations in regions where the depth should remain constant. As an example, the depth for the blue plastic container in figure 1(b)(ii) is heavily distorted. In figure 1(c)(ii), there are significant depth variations in the white lid of the Centrum pill bottle and along the bottle edges. Likewise, in figure 1(d)(ii), similar issues are witnessed on the cardboard figurine and in the background. The distance speckling is due, in part, to the low amount of texture variation in these regions across each focal-stack image. It is also a byproduct of having only a few, low-resolution focal stack images.

Our variational framework was largely able to avoid producing depth maps with such errors. Moreover, there appear to be few, visually evident, drawbacks to using regularization on these scenes. For example, the basic shape of the pill bottle in figures 1(c)(iv)–(vii) remains the same regardless of the regularization amount. Only sporadic fluctuations become increasingly suppressed as the weight is raised. Much about the shape of the blue plastic container in figure 1(b)(i) and the cardboard figurine in figure 1(d)(i) is retained in figures 1(b)(iv)–(vii) and 1(d)(iv)–(vii) as the weight is increased. It is important to note that these results could not have been obtained by simply post-processing the results from the sum-modified Laplacian model. In figure 1(b)(ii), for instance, there are significant distance variations in the plastic container that could not be easily removed without also significantly perturbing the remaining scene depth. Applying averaging and median filters or morphological operations to the remaining scenes removed many important depth structures and led to some distance bleeding across various objects.

For figures 1(b) and 1(d), learning a metric versus not doing so yielded better qualitative depth maps. The distance mottling was reduced for these scenes, while the object shape was better recovered. In figures 1(a) and 1(c), not much improvement was achieved. This was largely because corresponding scenes had the highest number of focal images. The high-frequency texture components could be better discerned by the contrast measure, which led to local minima that were closer to the global minima than in the remaining two scenes.

There are at least two reasons as to why we obtained such promising depth maps. First, our framework is designed to simultaneously consider both the sharpness measure and the smoothness constraint across all focal images when assigning a depth value. This behavior allows the framework to entertain multiple depth hypotheses when making a decision about object distance. Second, the total variation regularizer is adept at removing localized variations and retaining edges that correspond to widespread jumps in depth.

There are some general comments that can be made about the results. One concerns the lack of detailed depth variation in the backgrounds for many of these scenes. For example, in figure 1(b), some of the books, pens, and papers are treated as a single object and assigned a near-consistent distance. In figure 1(b), there are a number of stacked boxes in the focal stack that are given a near-constant depth value. These responses are caused by the limited number focal plane locations and hence focal images. To capture the depth variations of these background objects, more focal images would be needed.

Another comment concerns the applicability of our method to objects with certain visual properties. In figure 1(c), the depth outline of the pill bottle is rather jagged compared to the statue outline in figure 1(a). There is also some ambiguity about the shape of the blue plastic container in figure 1(b). These issues are caused by specular reflections near the curved container edges. These reflections give the edges a faintly hazed appearance across the focal images. Consequently, there are few high-frequency components, and it becomes difficult to gauge when the edges of the containers are in focus. Mirrored objects pose even more difficulties: the sum-modified Laplacian measure falsely concludes that these objects have significant variations in depth, as in figure 1(d). Most, if not all, existing depth-from-focus methods also suffer from the same issues, though.

REFERENCES

[1] R. A. Jarvis, "A perspective on range finding techniques for computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 122–139, 1983. Available: http://dx.doi.org/10.1109/TPAMI.1983.4767365

[2] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 223–237, 1988. Available: http://dx.doi.org/10.1007/BF00127822

[3] M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," *International Journal of Computer Vision*,

vol. 27, no. 3, pp. 203–225, 1999. Available: http://dx.doi.org/10.1023/A:1007905828438

[4] A. N. Rajagopalan and S. Chaudhuri, "An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 577–589, 1999. Available: http://dx.doi.org/10.1109/34.777369

[5] P. Favaro, S. Soatto, M. Burger, and S. J. Osher, "Shape from defocus via diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 518–531, 2008. Available: http://dx.doi.org/10.1109/TPAMI.2007.1175

[6] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994. Available: http://dx.doi.org/10.1109/34.308479

[7] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 97–108, 1993. Available: http://dx.doi.org/10.1109/34.192482

[8] S. Das and N. Ahuja, "Performance analysis of stereo, vergence, and focus as depth cues for active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1213–1219, 1995. Available: http://dx.doi.org/10.1109/34.476513

[9] M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 864–870, 1998. Available: http://dx.doi.org/10.1109/34.709612

[10] N. Asada, H. Fujiwara, and T. Matsuyama, "Edge and depth from focus," *International Journal of Computer Vision*, vol. 26, no. 2, pp. 153–163, 1998. Available: http://dx.doi.org/10.1023/A:1007996810301

[11] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition*, vol. 46, no. 5, pp. 1415–1432, 2013. Available: http://dx.doi.org/10.1016/j.patcog.2012.11.011

[12] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 266–274, 1995. Available: http://dx.doi.org/10.1109/34.368191

[13] M. T. Mahmood and T.-S. Choi, "Nonlinear approach for enhancement of image focus volume in shape from focus," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2866–2873, 2012. Available: http://dx.doi.org/10.1109/TIP.2012.2186144

[14] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, 1987. Available: http://dx.doi.org/10.1109/TPAMI.1987.4767940

[15] S.-H. Lai, C.-W. Fu, and S. Chang, "A generalized depth estimation algorithm with a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 405–411, 1992. Available: http://dx.doi.org/10.1109/34.126803

[16] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, 2007. Available: http://dx.doi.org/10.1007/BF02028349

[17] M. Muhammed, H. Mutahira, A. Majid, and T.-S. Choi, "Recovering 3D shape of weak textured surfaces," in *Proceedings of the IEEE International Conference on Computational Science*, Yongin, South Korea, June 29-July 2 2009, pp. 191–197. Available: http://dx.doi.org/10.1109/ICCSA.2009.25

[18] S. Pertuz, D. Puig, and M. A. Garcia, "Reliability measure for shape-from-focus," *Image and Vision Computing*, vol. 31, no. 10, pp. 725–734, 2013. Available: http://dx.doi.org/10.1016/j.imavis.2013.07.005

[19] V. Gaganov and A. Ignatenko, "Robust shape from focus via Markov random fields," in *Proceedings of the International Conference on Computer Graphics and Vision*, Moscow, Russia, October 5-9 2009, pp. 74–80.

[20] A. S. Malik and T.-S. Choi, "Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery," *Pattern Recognition*, vol. 40, no. 1, pp. 154–170, 2007. Available: http://dx.doi.org/10.1016/j.patcog.2006.05.032

[21] A. Thelen, S. Frey, S. Hirsch, and P. Hering, "Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood size, and height value interpolation," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 151–157, 2009. Available: http://dx.doi.org/10.1109/TIP.2008.2007049

[22] C. L. Byrne, *Iterative Optimization in Inverse Problems*. Boca Raton, FL, USA: CRC Press, 2014.

[23] C. Udriste, *Convex Functions and Optimization Algorithms on Riemannian Manifolds*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1994.

[24] S. Hosseini and M. R. Pouryayevali, "Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds," *Nonlinear Analysis*, vol. 74, no. 12, pp. 3884–3895, 2011. Available: http://dx.doi.org/10.1016/j.na.2011.02.023

[25] F. H. Clarke, "Generalized gradients and applications," *Transactions of the American Mathematical Society*, vol. 205, no. 1, pp. 247–262, 1975. Available: http://dx.doi.org/10.2307/1997202