Interactive Control Approach to 3D Shape Reconstruction

Bipul Islam, Ji Liu, Anthony Yezzi, Romeil Sandhu

***Abstract*— The ability to accurately reconstruct the 3D facets of a scene is one of the key problems in robotic vision. However, even with recent advances with machine learning, there is no high-fidelity universal 3D reconstruction method for this optimization problem as schemes often cater to specific image modalities and are often biased by scene abnormalities. Simply put, there always remains an "information" gap due to the dynamic nature of real-world scenarios. To this end, we demonstrate a feedback control framework which invokes operator inputs (also prone to errors) in order to augment existing reconstruction schemes. For proof-of-concept, we choose a classical region-based stereoscopic reconstruction approach and show how an ill-posed model can be augmented with operator input to be much more robust to scene artifacts. We provide necessary conditions for stability via Lyapunov analysis and perhaps more importantly, we show that the stability depends on a notion of absolute curvature. Mathematically, this aligns with previous work that has shown Ricci curvature as proxy for functional robustness of dynamical networked systems. We conclude with results that show how our method can improve standalone reconstruction schemes.**

I. Introduction

Sensing the spatial particulars and inferring information about a real-world scene from images is a classical problem in robotic vision with a multitude of uses ranging from motion planning, situational awareness, to medical imaging [1], [2], [3]. This said, reconstruction of a complex 3D scene from 2D images is a difficult task due to the amount of uncertainties that must be accounted for in real-world scenarios. Although much progress have been made over the last few decades, reconstruction methodologies often fail as a result of imaging artifacts including, but not limited to, noise, occlusions, clutter, and non-uniform illumination. In short, no universal algorithm exists which can work seamlessly across all image modalities [4]. To combat such risk complexities, there is a need for domain experts or an operator who is able to provide an estimate of the ideal result and subsequently able to verify the quality of reconstruction. Here, we aim to "inject" 2D operator inputs in-loop to drive a (multi-agent) 3D surface deformation while ensuring the resulting system is stable in the sense of Lyapunov [5]. While this work builds off of our previous work in image segmentation [4] and reconstruction [1], there lies a few tacit yet important discerning caveats. **Firstly**, we show that 2D operator inputs of a given set of images can be aptly "mapped" to 3D world and such inputs, are stable. Mathematically, this not a trivial issue as any input on a 2D background should also be corroborated by a 3D action on infinitely large ("blue sky") background (e.g., specifying the 3D action location based on 2D background input is ill-posed). From a stability perspective, such singular 2D actions affect not only a 3D surface deformation, but indirectly affect other 2D passive sensors via 3D-to-2D projections during the reconstruction process. **Secondly**, the control laws are developed in-part based on a notion of absolute principle curvature which is a main underlying theme of this work (e.g., confluence of geometry & control). **Thirdly**, curvature can be shown to relate to a notion of "trust" in the sense of how quickly our reconciled solution converges from both the operator and autonomous perspective. This will be stylized in detail in future work, but is presented here to place this work and contributions in context. We now briefly revisit a few techniques as it pertains to this work.

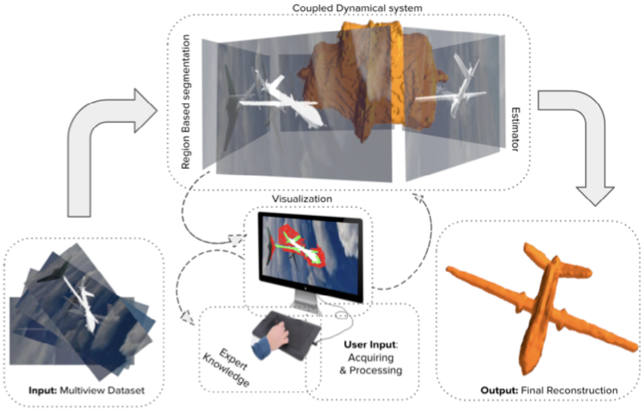

Coupled Dynamical system
Region Based segmentation
Estimator
Visualization
Expert Knowledge
User Input: Acquiring & Processing
Input: Multiview Dataset
Output: Final Reconstruction

Fig. 1: Schematic outline of interactive feedback control stereoscopic reconstruction framework.

A. Brief 3D Reconstruction Literature Review

Most modern scene reconstruction methods use the popular deep (reinforcement) learning variants and are often characterized by the requirement of massive training samples [6], [7]. Some examples of such systems are ScanNet [6] that uses over 2.5 million scenes to train a system that can understand indoor scenes to [7] where authors furnish a synthetic dataset in order to develop an understanding of surface normal prediction, semantic segmentation, and object boundary detection. Generally, such schemes are highly dependent on the training quality. To combat this, [8] explores the use of supervision as an alternative for expensive 3D annotation from which perspective projection and back propagation are employed. On the other hand, such methods

B. Islam and R. Sandhu are with the Department of Computer Science & Biomedical Informatics and J. Liu is with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794. A. Yezzi is with the Department of Electrical and Computer Engineering, Georgia Tech, Atlanta GA, 30309. R. Sandhu is also with the Departments of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY 11794. E-mail: bipul.islam@stonybrook.edu

use local correspondence matching and hence, are fallible to drawbacks resulting from scene abnormalities (e.g., noise, non-uniform illumination [9]). In regards to robotic vision, such correspondence-based solutions generally involve the well-known concept of SLAM (Simultaneous Localization and Mapping) [10], [11], [12]. This said, SLAM-based methods traditionally suffer from the requirement of high computational power for sensing a sizable area and process the resulting data to perform both mapping and localization. Also, there is a tacit requirement that input scene images should have overlap from image-to-image. To this end, SFM (Structure From Motion) based methods provide a relaxed version of this problem [13], [14] (i.e., Google uses this approach in their popular street-view application on Google maps [15]). More recently, [16] explores a recurrent neural network (3D-R2N2) by employing shape priors in which one learns 2D to 3D mapping from images of objects to their underlying 3D shapes from large collections of synthetic data. In particular, the authors have been seemingly able to show their method outperforming SLAM or SFM (albiet with learnt knowledge) when there is lack of texture or baseline.

Nevertheless, this paper does not argue the rigors of the underlying reconstruction method itself and our particular focus on our previous work [1] is in-part due a correspondence-free method, independence to local (image-gradient) structure, and dependence on geometric techniques connected to image segmentation [17], [21], [22]. Undoubtedly, each approach whether it be SLAM-based, deep (reinforcement) learning variants, and/or geometric methods work optimally with respect to the prospective operating environment (e.g., space, low-power requirements compared ground-based robotic vision). At the same time, any such reconstruction are not infallible to errors that arise in real-world dynamic scenes from a human-perception standpoint. This said, human-perception is also fallible and any operator input based on a visual estimate is prone to errors. Philosophically, we make the argument that terms such as over-fitting and uncertainty are in part, perceived by an expert who generally acts as a passive entity in such methods. Thus, the problem we seek to resolve is to not only rectify the expected and ideal reconstruction in real-time [23], but provide the necessary feedback control characterization when invoking operator input [24].

The remainder of the paper is organized as follows: In the next, we introduce stereoscopic reconstruction via classic image segmentation. Then Section III provides a control framework along with the necessary conditions for stability. Section IV presents experimental results. From this, we conclude with future work in Section V.

II. From Segmentation to 3D Reconstruction

This section presents a general introduction to geometric stereoscopic segmentation.

A. Geometric 2D Image Segmentation

Let us begin with the classic binary problem of segmenting an image $I : \Omega \mapsto \mathbb{R}^n$ into a foreground and background described by functionals $r_o : \zeta,\ \Omega \mapsto \mathbb{R}$ and $r_b : \zeta,\ \Omega \mapsto \mathbb{R}$ which measure the similarity of of the image pixels with a statistical model over the regions R and R^c, respectively. Here, ζ corresponds to the photometric variable of interest. Then, one can define a partitioning problem where the optimal partition between foreground/background is described by a partial differential equation [22], [26]; i.e.,

$$\begin{aligned} E &= \int_R r_o(I(\mathbf{x}),C) + \int_{R^c} r_b(I(\mathbf{x}),C) d\Omega \qquad (1) \\ \frac{\partial E}{\partial C} &= \beta \vec{N} \end{aligned}$$

where $\beta : \mathbb{R}^2 \mapsto \mathbb{R}$ can be considered "forces" along the curve (partition boundary) that describe the direction of the corresponding evolution in the normal $\vec{N}$ direction. While a complete review of such methodology is beyond the scope of this note, we do refer the reader to several seminal references [17], [21]. For the case image segmentation, it suffices to understand that the partitioning curve C "lives" in the 2D image domain.

B. Stereoscopic 3D Reconstruction

Now, if we consider the problem of 3D reconstruction from 2D images, one can redefine the functional in equation (1) as follows:

$$E = \sum_{i=0}^{N} \int_{R_i} r_o(I_i(\hat{\mathbf{x}}_\mathbf{i}), \pi_i^{-1}(\hat{\mathbf{x}}_\mathbf{i}), \hat{c}_i) + \int_{R_i^c} r_b(I_i(\hat{\mathbf{x}}_\mathbf{i}), \Theta_i(\hat{\mathbf{x}}_\mathbf{i}), \hat{c}_i) d\Omega_i \qquad (2)$$

where the difference is the functional now depends on N image observations I_i and where a particular 2D image silhouette curve $\hat{c}_i$ is derived from a single 3D occluding curve C (with a slight abuse of notion) on a given smooth surface S in $\mathbb{R}^3$ with a corresponding 3D background B treated as infinitely large sphere with angular coordinates $\Theta = (\gamma, \upsilon)$. That is, $\hat{c}_i = \pi_i(C)$ where $\pi_i : \mathbb{R}^3 \mapsto \Omega_i$ is the realization of the i-th pin-hole camera (sensor) that projects the 3D world onto the 2D domain. Similarly, the background can be related in a one-to-one manner with the image coordinates $\hat{\mathbf{x}}_\mathbf{i}$ of each observation through the mapping Θ_i ("blue sky" assumption). To be more precise, $\mathbf{x} = (x,y,z)$ is surface coordinates of S in $\mathbb{R}^3$ and further note that $\mathbf{x}_\mathbf{i} = (x_i, y_i, z_i)$ denote the same points expressed in i-th calibrated camera coordinates relative to the i-th image. Moreover, $\hat{\mathbf{x}}_\mathbf{i} = (\hat{x}_i,\ \hat{y}_i) = (x_i/z_i, y_i/z_i)$ is the aforementioned perspective projection due to the i-th pin-hole camera π_i. In turn, r_o and r_b redefined to be radiance functions. That is, the foreground object of interest supports a radiance function of r_o: $S \rightarrow \mathbb{R}$ with the usual area element dA. Similarly, the background supports a different radiance function r_b: $B \rightarrow \mathbb{R}$. As such, for a given 3D surface, it is possible to partition each image domain Ω_i of I_i into a foreground object region $R_i = \pi_i(S) \subseteq \Omega_i$ and the corresponding background region R_i^c. Note, the operator π_i is not one-to-one and, hence non-invertible. However, we can define a back projection operator π_i^{-1} using the back tracing of rays from image to the surface, i.e, we have $\pi_i^{-1} : R_i \rightarrow S$ which is a pseudo one-to-one operation.

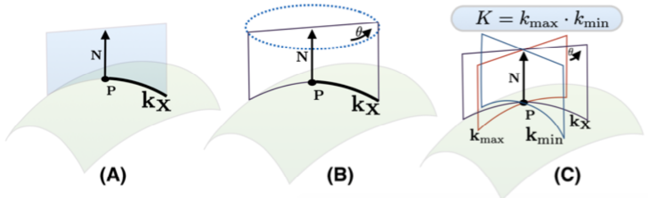


Fig. 2: Visualization of Normal, Principle, and Gaussian Curvature. (A) Given Point P and Normal N, define a perpendicular plane intersection at point P. The curve that plane intersects on the manifold is known as the normal curvature κ_X in the direction X. (B) We can define other normal curvatures through a rotation of the plane by θ. (C) The max and min normal curvature are what is known as principle curvatures. The product of these principle curvatures yields Gaussian curvature.

Putting this together, assuming the calibrated cameras, the deformation of the surface towards a reconstructed shape based on a set of N image observations can be shown to be of the following form:

$$\frac{\partial S}{\partial t} = \sum_{i=0}^{N} \beta_i \cdot \left(\frac{\nabla_{x_i} \chi_i \cdot \mathbf{x_i}}{z_i^3} \right) \vec{N}$$

where we define a visibility characteristic function χ_i from a given location $\mathbf{x_i}$ on a surface S as:

$$\chi_i(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \mathbf{x_i} \in \pi_i^{-1}(R_i) \\ 0, & \text{if } \mathbf{x_i} \notin \pi_i^{-1}(R_i). \end{cases}$$

This can be re-written in terms of the smooth regularized-Heaviside function H along with (outward) surface normals $\vec{N}$ at each point $\mathbf{x_i}$ of the surface S:

$$\chi_i = 1 - H(\mathbf{x_i} \cdot \vec{N})$$

Given the above, we are now able to formulate a control-based reconstruction scheme from which a given physical 2D action, based on visual perception (information), can be used to interactively "sculpt" a 3D shape in collaboration with the above autonomous 3D reconstruction algorithm.

III. Control-Based Reconstruction

Let us begin by redefining the general form of a surface reconstruction evolution above in level-set notation as follows:

$$\frac{d\phi}{dt} = \sum_{i=0}^{N} \psi_i(\hat{\mathbf{x}}_\mathbf{i}, \mathbf{x_i}, t)\delta(\phi(\mathbf{x})) \tag{3}$$

where $\psi_i : \mathbb{R}^3 \to \mathbb{R}$ is the surface gradient information computed from the photometric image data, $\phi : \mathbb{R}^3 \to \mathbb{R}$ is a level-set function, and $\delta(.)$ is the classical Kronecker delta function. Hence, to "close the loop" that incorporates a physical 3D operator performing 2D inputs in order to control the 3D evolution dynamics of the evolving surface, one has

$$\frac{d\phi}{dt} = \sum_{i=0}^{N} [\psi_i + F_i(\phi, \phi^*)]\delta(\phi) \tag{4}$$

where F_i is the to be defined control law that drives ϕ towards the ideal (perfect) surface ϕ^* as $t \to \infty$. The definition of an ideal surface is this note is a result with no errors. For this work, we use the mean-separable segmentation energy [21] as our reconstruction model. From this, $\nabla_{x_i} \chi_i \cdot \mathbf{x_i}$ can be expressed in terms of curvature for points on the surface which leads us to the following Lemma.

Lemma III.1 *For a given characteristic function χ_i and a point $\mathbf{x_i} \in S$ that lies on the corresponding surface "imaged" from a given camera π_i, we have that*

$$\nabla_{x_i} \chi_i \cdot \mathbf{x_i} = -\kappa_u \, \|\mathbf{x_i}\|^2 \, \delta(\mathbf{x_i} \cdot \vec{N}). \tag{5}$$

Proof: Following the nomenclature defined above and noting $\mathrm{II}(\boldsymbol{x}, \boldsymbol{x})$ is the second fundamental form [14], [27], we have

$$\begin{aligned} \nabla_{x_i} \chi_i \cdot \mathbf{x_i} &= \langle \nabla_{x_i} (1 - H(\mathbf{x_i} \cdot \vec{N})), \mathbf{x_i} \rangle \\ &= -\langle \delta(\mathbf{x_i} \cdot \vec{N}) \nabla_{x_i} (\mathbf{x_i} \cdot \vec{N}), \mathbf{x_i} \rangle \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) \langle \nabla_{x_i} (\mathbf{x_i} \cdot \vec{N}), \mathbf{x_i} \rangle \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) (\nabla_{x_i} \vec{N}^T \mathbf{x_i})^T \mathbf{x_i} \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) [\mathbf{x_i}^T \nabla_x \vec{N} \mathbf{x_i}] \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} l & m \\ m & n \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) \; \mathrm{II}(\mathbf{x_i}, \mathbf{x_i}) \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) \frac{\mathrm{II}(\mathbf{x_i}, \mathbf{x_i})}{\mathbf{x_i}^T \mathbf{x_i}} \, \|\mathbf{x_i}\|^2 \\ &= -\delta(\mathbf{x_i} \cdot \vec{N}) \kappa_u \, \|\mathbf{x_i}\|^2 \end{aligned} \tag{6}$$

where k_u is the **normal curvature** *in a particular viewing direction* $\mathbf{x_i}$ on the corresponding surface S. We refer to Figure 2 for a visualization of this type of curvature on a given manifold. From this, we can rewrite ψ_i as the following:

$$\psi_i = -\beta_i \frac{\delta(\mathbf{x_i} \cdot \vec{N}) \kappa_u \, \|\mathbf{x_i}\|^2}{z_i^3}. \tag{7}$$

Furthermore, as we aim to define a control law F_i such that $\lim_{t \to \infty} \phi(\boldsymbol{x}) \to \phi^*(\boldsymbol{x})$, we define the error between our current estimate and ideal shape (no errors) as

$$E_e(\mathbf{x}, t) := H(\phi(\mathbf{x}, t)) - H(\phi^*(\mathbf{x})). \tag{8}$$

In doing so, we are now able to define the existence of the control law F_i via Lyapunov method of stabilization.

Theorem III.1: *Let us assume $z_i \geq 1$ and $||\mathbf{x_i}||^2 \leq z_i^3$ as well as let κ_{max} and κ_{min} be the the* ***principle maximum curvature*** *and* ***principle minimum curvature*** *at a given point $\mathbf{x_i}$ with respect to an imaging referential camera π_i, respectively. Then the control law*

$$F_i = - \mid \beta_i \mid \kappa_{abs} E_e \tag{9}$$

where $\kappa_{abs} = \mid \kappa_{min} \mid + \mid \kappa_{max} \mid$, asymptotically stabilizes the system given in equation (4) from the current evolving surface $\phi(\boldsymbol{x}, t)$ to the ideal surface, $\phi^*(\boldsymbol{x})$ as $t \to \infty$.

Proof: We choose the Lyapunov function $V(E_e,t) \in C^1$ defined in terms of $E_e(\mathbf{x},t)$ as

$$V = \frac{1}{2}\int_{S\cup S^*} \|E_e(\mathbf{x},t)\|^2 dx. \qquad (10)$$

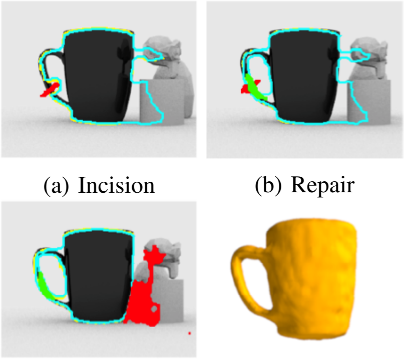
Differentiating V with respect to time t we get:

$$\begin{aligned}\frac{\partial V}{\partial t} &= \int_{S\cup S^*} E_e \frac{\partial E_e}{\partial t} dx \\ &= \int_S E_e[\delta(\phi)\frac{\partial \phi}{\partial t}]dx \\ &= \int_S E_e \delta(\phi)[\sum_{i=0}^{N}[\psi_i + F_i]\delta(\phi)]dx\end{aligned} \qquad (11)$$

The simplification over the union $S \cup S^*$ results from the application of the Kronecker delta function. Moreover, one can show that resulting system is stable (i.e., V has a negative semidefinite derivative):

$$\begin{aligned}\frac{\partial V}{\partial t} &= \sum_{i=0}^{N}\int_S E_e\delta(\phi)^2[\psi_i + F_i]dA \\ &= \sum_{i=0}^{N}\int_S E_e\delta(\phi)^2\left[\beta_i \cdot \frac{\nabla_{x_i}\chi_i \cdot \mathbf{x_i}}{z_i^3} + F_i\right]dA \\ &= \sum_{i=0}^{N}\int_S E_e\delta(\phi)^2\left[\beta_i \cdot \frac{\nabla_{x_i}\chi_i \cdot \mathbf{x_i}}{z_i^3} - \mid \beta_i \mid \kappa_{abs}E_e\right]dA \\ &= \sum_{i=0}^{N}\int_S \delta(\phi)^2\left[E_e \cdot \beta_i \cdot \frac{\nabla_{x_i}\chi_i \cdot \mathbf{x_i}}{z_i^3} - \mid \beta_i \mid \kappa_{abs}E_e^2\right]dA \\ &\leq \sum_{i=0}^{N}\int_S \delta(\phi)^2\left[E_e^2 \cdot \mid \beta_i \mid \left|\frac{\nabla_{x_i}\chi_i \cdot \mathbf{x_i}}{z_i^3}\right| - \mid \beta_i \mid \kappa_{abs}E_e^2\right]dA \\ &= \sum_{i=0}^{N}\int_S \delta(\phi)^2 E_e^2 \mid \beta_i \mid \left[\left|\frac{\nabla_{x_i}\chi_i \cdot \mathbf{x_i}}{z_i^3}\right| - \kappa_{abs}\right]dA \\ &= \sum_{i=0}^{N}\int_S \delta(\phi)^2 E_e^2 \mid \beta_i \mid \left[\mid \kappa_u \frac{\|\mathbf{x_i}\|^2}{z^3} \mid - \kappa_{abs}\right]dA \\ &< \sum_{i=0}^{N}\int_S \delta(\phi)^2 E_e^2 \mid \beta_i \mid [\mid \kappa_u \mid - \kappa_{abs}]dA \\ &\leq 0\end{aligned}$$

In particular, the above control law will be dependent on curvature. While beyond the scope of this note, one can show exponential convergence whereby higher curvature coincides with faster convergence rates. While we have not included this derivation in the present work due to scope and for sake of clarity, we will expound upon this in future work. This said, we present such comments to better highlight important caveats in terms of geometry and control as well as how one can start to define notions of "trust" (from a reconciliation of an operator augmentation) to that of a geometric (curvature) quantity. We would like to highlight there exists analogous behavior in networked dynamical systems in which one is able to use discrete Ricci curvature as a measure for network robustness [28]. In such work, one can leverage the concept of k-convexity similarly to above to define positive correlation between Boltzmann entropy, curvature, and rate functions from thermodynamics. Ultimately, this work will seek to build upon this area and in particular, explore notions of "trust" in the sense of geometric quantities such as curvature.



(a) Incision (b) Repair

(c) Consolidate (d) Final

Fig. 3: A summary of operators actions to maneuver out of the local minima in a complex occluded scene. The images (A), (B), and (C) are views of the model after each of interaction milestones. Sub-figure (C) shows the final reconstruction.

Nevertheless, in designing operator guided inputs, we note perfect knowledge of ideal surface is not readily available (even from a human visualization perspective) due a myriad of reasons including, but not limited to, occlusions, clutter, and/or inability to define a well-posed model across image modalities. As such, we allow an operator (whom is also prone to errors) to make interactions with the system in order to reconcile *one's belief* with built autonomy towards an estimate of the ideal surface. We stress the fact that the input from a human is fallible and such input indirectly affects our control law through the adjudication of an "ideal" estimate. This estimate herein is denoted as $\hat{\phi}^*(\mathbf{x},t)$. Moreover, we define $\varepsilon_i^k(\hat{\mathbf{x}}_\mathbf{i},t)$ as the k-th input on a given image i and the accumulated input $U_i : \mathbb{R}^2 \to \mathbb{R}$ as

$$\begin{aligned}\varepsilon_i^k(\hat{\mathbf{x}}_\mathbf{i},t) &:= \pm p \text{ (constant)} \\ U_i(\hat{\mathbf{x}}_\mathbf{i},t) &:= \sum_{l=0}^{k}\varepsilon_t^k(\hat{\mathbf{x}}_\mathbf{i}).\end{aligned}$$

That is, we seek to allow for *the physical operator to make 2D actions such that it will deform a 3D surface*. In other words, we are able to define a 3D control law based on 2D inputs which is particularly helpful as the operator is generally ill-equipped to alter the 3D shape itself (i.e., we assume the operator not to be an artist). To derive the coupled system that fuses 2D operator input to control the 3D surface deformation, we must also define the errors for **both** the operator and autonomous model:

$$\begin{aligned}E_A &= H(\phi(\mathbf{x}),t)) - H(\hat{\phi}^*(\mathbf{x}),t)) \\ E_{u_i} &= H(\hat{\phi}^*(\boldsymbol{x}),t)) - H(U_i(\pi_i(\boldsymbol{x}),t)).\end{aligned} \qquad (12)$$

Given this, we can now define a coupled PDE system that unifies both the operator based inputs along with that of the autonomous counterpart which is representative of an estimator-observer behavior as follows:

$$\frac{\partial \phi}{\partial t} = \sum_{i=1}^{N} [\psi_i + F_i]\delta(\phi) \tag{13a}$$

$$\phi(x,0) = \phi^0(x)$$

$$\frac{\partial \hat{\phi}^*}{\partial t} = \sum_{i=1}^{N} [E_A + f_i(U_i, E_{u_i})] \tag{13b}$$

$$\hat{\phi}^*(x,0) = \phi^0(x)$$

where the tuning function $f_i(U_i, E_{u_i})$ that is dependent on operator input from an image observation can be defined as

$$f_i(U_i, E_{u_i}) = - \mid U_i \mid E_{u_i}. \tag{14}$$

This said, the above system then needs to be shown that it is is still stable even from imperfect operator actions. To do so, we define the accumulated total errors for both the operator and autonomous model as

$$E(t) := \frac{1}{2}\sum_{i=1}^{N} \int_{S \cup B} \mid U_i \mid E_{u_i}^2 dx \tag{15}$$

$$\Gamma(t) := \frac{1}{2}\sum_{i=1}^{N} \int_{S \cup S^*} E_A^2. \tag{16}$$

From this, we now arrive at the following result.

Theorem III.2: *Let us assume previous notation and results in Theorem III.1 and further assume that operator input has stopped (i.e., U_i is constant in all viewing directions), then the estimator*

$$\frac{\partial \hat{\phi}^*}{\partial t} = \sum_{i=1}^{N} [E_A + f_i(U_i, E_{u_i})]$$

where $f_i(U_i, E_{u_i}) = - \mid U_i \mid E_{u_i}$ will stabilize the resulting coupled system in equation (13a) and equation (13b). Namely, the total error $\Phi(t) := E(t) + \Gamma(t)$ has a negative semidefinite derivative.

Proof. Let us begin by differentiating $E(t)$ with respect to t:

$$\begin{aligned} \frac{\partial E}{\partial t} &= \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \mid U_i \mid E_{u_i} \frac{\partial E_{u_i}}{\partial t} \\ &= \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \mid U_i \mid E_{u_i} \delta(\hat{\phi}^*) \frac{\partial \hat{\phi}^*}{\partial t} dx \\ &= \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \mid U_i \mid E_{u_i} \delta(\hat{\phi}^*)[E_A - \mid U_i \mid E_{u_i}] dx. \end{aligned} \tag{17}$$

Similarly, differentiating $\Gamma(t)$ with respect to t:

$$\begin{aligned} \frac{\partial \Gamma}{\partial t} &= \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} E_A \left[\delta(\phi) \frac{\partial \phi}{\partial t} - \delta(\hat{\phi}^*) \frac{\partial \hat{\phi}^*}{\partial t} \right] dx \\ &= \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \delta(\phi)^2 E_A [\psi_i + F_i] dx \\ &\quad - \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \delta(\hat{\phi}^*)^2 E_A [E_A - \mid U_i \mid E_{u_i}] dx. \end{aligned} \tag{18}$$

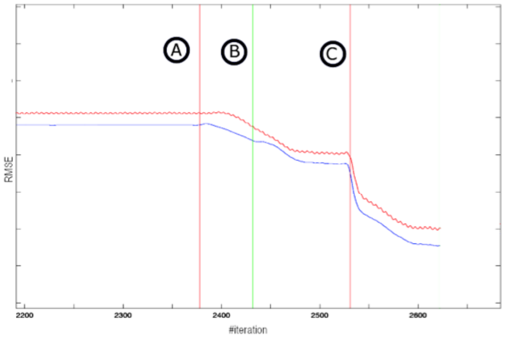


Fig. 4: The overall sequence w.r.t. to energy minimization of operator action corresponding to Figure 3b. (A) Incision, (B) Repair, (C) Consolidate.

From this, we are now able to combine terms for the total labeling error $\Phi(t) = E(t) + \Gamma(t)$. That is, summing equation (17) and equation (18) and simplifying:

$$\begin{aligned} \frac{\partial \Phi}{\partial t} &= \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \mid U_i \mid E_{u_i} \delta(\hat{\phi}^*)[E_A - \mid U_i \mid E_{u_i}] dx \\ &+ \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \delta(\phi)^2 E_A [\psi_i + F_i] dx \\ &- \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \delta(\hat{\phi}^*)^2 E_A [E_A - \mid U_i \mid E_{u_i}] dx \\ &\leq - \sum_{i=1}^{N} \int_{S \cup \hat{S}^*} \delta(\hat{\phi}^*)^2 [E_A - \mid U_i \mid E_{u_i}]^2 dx \\ &\leq 0 \end{aligned} \tag{19}$$

As the derivative is negative semidefinite, the coupled system defined above is stable. □

IV. Experimental Results and Discussion

In this section, we demonstrate the proposed algorithm on a variety of scenarios. In all demonstrated results, green patches, or marks, are made by the user to denote regions in the foreground. Similarly, red denotes regions on images that are to be considered a part of the background. In images where silhouettes are displayed, the yellow silhouette denotes the autonomous surface while the estimate of ideal surface is always presented in cyan. Each reconstruction utilizes $N = 36$ images with the resulting MATLAB code run on an iMac 4.2 Ghz Core i7 with 32GB memory.

We begin with an example that highlights the method in face of **occlusions** by objects obfuscating several different imaging views. This can be seen Figure 3 along with how such inputs affect the energy minimization landscape in Figure 4. Here, naive reconstruction fails due to ambient occlusion whose intensity is similar to the background. While there exists varying approaches and shape prior models to overcome such a problem, defining such models for particular scenarios becomes quite cumbersome and yet, may not yield stable results. We are able to properly reconstruct the shape through operator input with a simplified model as defined in [21]. For this experiment, the user made 12

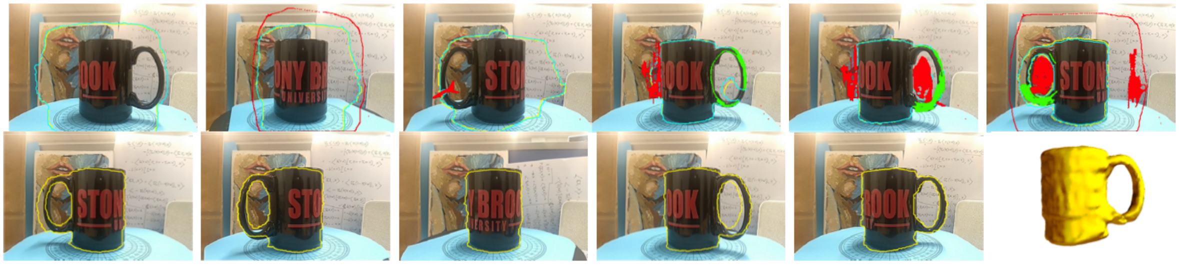


Fig. 5: 3D reconstruction of a cup in clutter and camera miscalibration. Top row: Sequence of user initiated operations to reorient the flow at multiple time instances. Bottom row: Final silhouette curves and reconstruction. Note: Yellow Curve is Autonomous Surface, Blue Curve is Ideal Estimate, Green is Foreground Interaction, Red is Background Interaction.

interactions for the foreground and 47 interactions for the background. In particular, in regards to the operator input and its impact on the energy landscape, the user actions can be partitioned into 3 milestones: initial incision (Figure: 3a), followed by a repair of the surface (Figure: 3b), and then, consolidating the surface by helping it "free" itself from scene anomalies (Figure: 3c).

More importantly, irrespective of the underlying model chosen for reconstruction, there will exist assumptions that are violated possibly due variety of image artifacts such as **noise, clutter, and/or model assumptions itself**. That is, for the chosen reconstruction autonomous model, we make the classic assumption that the scene is "mean-separable" and piecewise constant. Of course, while there exists other more advances models, such a model helps illustrate where operator feedback may override basic fallible assumptions. Figure 8 presents a scene in which such piecewise assumption is violated along minor camera miscalibrations. Additional scenes for which such assumption is violated can be seen in Figure 7 which aims to reconstructs a predator drone in a seemingly distinguishable background of clouds yet fails without operator input. In the context of stereoscopic reconstruction, overcoming non-uniform illumination is yet another tacit challenge. Figure 9 presents a scene where reconstruction of a sentinel drone fails due to tacit illumination on the ailerons that varies over the dataset. This is in part, due to illumination on the left wing which is consequently lower than the right wing. Utilizing operator input, the reconstruction results are demonstrated.

To further the idea in a **quantitive non-subjective manner**, we conduct numerical noise experiments on reconstruction of a synthetic scene of a sentinel drone which can be seen in Figure 6 and Table I. Ultimately, if the operator requires intensive work to assist the autonomous counterpart in such situations, then manual operator would suffice (or desire for improved built autonomy). This said, Table I presents **efficiency** results as the amount of user input is needed (in terms of % "actions" per view, % relabeling of pixels) compared to increased output (in terms of true and false positive rate pixel labels). For example, the second row can be stated that under 30% noise with only one action (user-input) on 95% of the views which amounts to only 2.7% pixel relabeling per image view, the true positive rate increases from 78.4% to 99.2%. This is repeated on several versions of noise and occlusion, two of which are seen from different views in Figure 6. *Nevertheless, the key application point of view here is that such failures of such reconstruction methods due to imaging artifacts such as noise can be naturally recover with minimal effort with human in-loop collaboration.* In addition to such results, we provide corresponding Lyapunov decay rates to such scenes in Figure 10.

Lastly, we note significant work on methods that use "feature"-based methods that rely on correspondences combined with machine learning to perform reconstruction tasks [6], [9], [8], [12]. While the thematic aspect of this paper is not discuss the rigors of such methods compared to the proposed underlying autonomous method, it is worthwhile to note that under such noisy situations, such correspondence methods (dependent on structural image information) began to suffer. Here, the geometric method proposed can be considered a "coarse" approach to tackle such "featureless" environments. This said, future work will focus on fusing such correspondence-based and learning approaches in hopes to define a notion of image integrity and leverage recent learning success on data that is indeed well-structured.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we have proposed a feedback control framework to guide the dynamics of an evolving surface in the context of multi-view stereoscopic reconstruction. This is done to ensure robustness in presence of low-fidelity datasets. From an optimization standpoint, the reconstruction

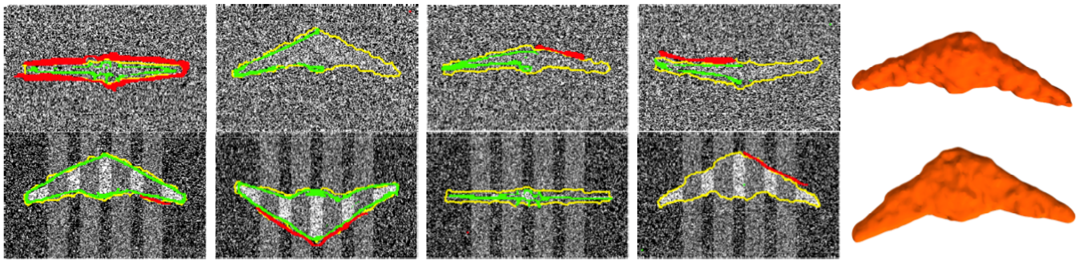


Fig. 6: Visual Synthetic Sentinel Drone Reconstruction Under Noise Conditions Corresponding to Table I. Top Row: Several Views Showing 90% Noise. Bottom Row: Several Views Showing 50% Noise with 37% Occlusion Over Image Domain.

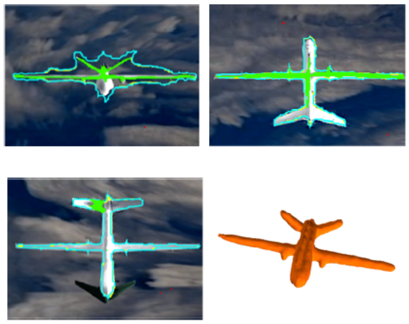


Fig. 7: Example where interactions are added to the wing-tips which are darker than ambient clutter of clouds and additional shape complication due drone thinness in wings.

Noise	Interactions % View	% Pixels	True Pos. Rate	False Pos. Rate
30%	(0, 0)	0%	78.4%	0.96%
30%	(1, 95%)	2.7%	99.2%	3.09%
50%	(0, 0)	0%	47.6%	0.2%
50%	(1, 95%)	2.7%	99.8%	4.8%
90%	(0, 0)	0%	21.8%	1%
90%	(1, 95%)	2.7%	99.9%	14.3%
90%	(1, 36%) (2, 60%)	6.3% (2.7+3.6)	99.9%	4.92%
noise: 50%+ occlusion: 37%	(1, 36%) (2, 60%)	6.7% (2.7+4)	99.6%	4.3%

TABLE I: Comparative analysis with noise and occlusion for the synthetic example of the Sentinel drone.

minima which we often seek (due to modeling imperfections) may not coincide with user expectations. As opposed to defining complex models for which overfitting may arise, we incorporate a user-defined input in-loop and "on-the-fly" from a feedback control perspective. We show the resulting framework is stable via Lyapunov analysis and from a practical standpoint, there is an increase in efficiency through a human-autonomous collaboration in shape reconstruction. Mathematically, the thematic interest is the interplay of geometry and control, namely how notions of curvature from geometry infer convergence and for this note, a notion of autonomous trust to user-input. This said, future work will entail a much closer analysis in regards to how Gaussian curvature infers convergence as well as the study of a problem in a distributed optimization sense, non-constant and time-delayed inputs as well as the inclusion of stochastic optimal control to further characterize operator uncertainty.

References

[1] A. Yezzi and S. Soatto. "Stereoscopic Segmentation." *International Journal of Computer Vision*. 2003

[2] O. Faugeras and R. Keriven. *Variational Principles, Surface evolution, PDE's, Level Set Methods and the Stereo problem*, *INRIA*. 1996.

[3] F. Zhao and X. Xie. "An Overview of Interactive Medical Image Segmentation", *Annals of the BMVA*. 2013

[4] L. Zhu, P. Karasev, I. Kolesov, I, R. Sandhu, and A. Tannenbaum. "Guiding Image Segmentation On The Fly: Interactive Segmentation From A Feedback Control Perspective", *IEEE Transactions on Automatic Control*. 2018.

[5] K. Khalil. "Nonlinear systems", *Prentice-Hall, New Jersey*. 1996.

[6] A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner. "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes", *CVPR*. 2017.

[7] Y. Zhang, S. Song, E. Yumer, M. Savva, J. Lee, H. Jin, and T. Funkhouser. "Physically-Based Rendering for Indoor Scene Understanding Using Convolutional Neural Networks", *CVPR*. 2017.

[8] J. Gwak. C. B. Choy, M. Chandraker, A. Garg, and S. Savarese. "Weakly Supervised 3D Reconstruction with Adversarial Constraint", *2017 International Conference on 3D Vision*. 2017.

[9] Z. Chen, X. Sun, L. Wang, Y. Yu and C. Huang. "A Deep Visual Correspondence Embedding Model for Stereo Matching Costs", *CVPR*. 2015.

[10] J. Aulinas, Y. Petillot, J. Salvi, X. Lladó. "The SLAM Problem: A Survey", *CCIA*. 2008.

[11] G. Zhang, and P. Vela. "Optimally Observable and Minimal Cardinality Monocular SLAM", *ICRA*. 2015.

[12] Y. Zhao and P. Vela. "Good Line Cutting: Towards Accurate Pose Tracking of Line-assisted VO/VSLAM", *ECCV*. 2018.

[13] A. Yezzi and S. Soatto. "Structure from Motion for Scenes without Features", *CVPR*. 2003.

[14] O. Faugeras. "Three-Dimensional Computer vision: a Geometric Viewpoint", *MIT Press*. 1993.

[15] B. Klingner, D. Martin, and J. Roseborough. "Street View Motion from Structure From Motion", *CVPR*. 2013.

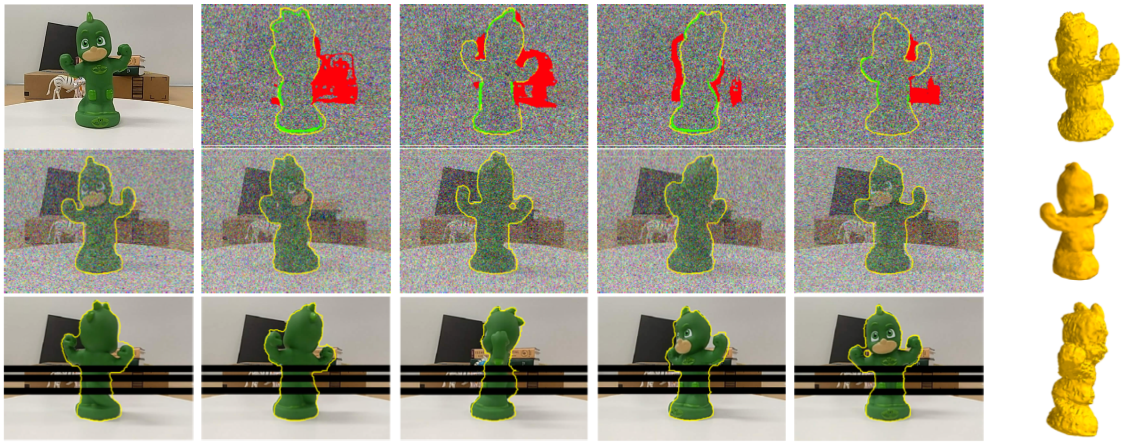
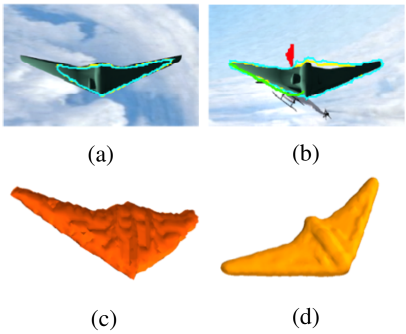


Fig. 8: 3D Reconstruction of a Toy Figurine. Top Row, First Image: Original Image of One View. Top Row: Significant Noise Applied to All Views with 3D Reconstruction. Middle Row: Moderate Noise Applied to All View with 3D Reconstruction. Bottom Row: Induced Camera Artifacts with 3D Reconstruction. Note: The underlying algorithm assumes object and background are "mean-separable" (e.g., such scenes are difficult corresponding to underlying autonomous model).



(a) (b)

(c) (d)

Fig. 9: Complications due to varied illumination conditions where interactions are added to the left wing-tip.

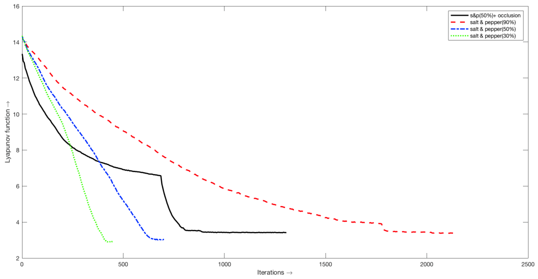


Fig. 10: Lyapunov function decay plots of noise scenes. Black: 50% Noise with Occlusion. Red: Noise at 90%. Blue: Noise at 50%. Green: Noise at 30%. The "knee-like" decrease in error in black/red signals are due to user-input.

[16] C. Choy, and D. Xu, J. Gwak, K. Chen, and S. Savarese. "3d-r2n2: A Unified Approach for Single and Multi-View 3D Object Reconstruction", *ECCV*. 2016

[17] D. Mumford and J. Shah. "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems", *Communications on Pure and Applied Mathematics*.1989.

[18] K. Kutulakos, N. Kiriakos, and S. Seitz. "A Theory of Shape by Space Carving", *International Journal of Computer Vision*. 2000.

[19] A. Mulayim, U. Yilmaz, and V. Atalay. "Silhouette-Based 3D Model Reconstruction From Multiple Images", *IEEE Transactions on Systems, Man, and Cybernetics*. 2003.

[20] M. Jancosek and T. Pajdla. "Segmentation Based Multi-View Stereo." *Computer Vision Winter Worskhop*. 2009.

[21] T. Chan and L. Vese. "An Active Contour Model Without Edges", *International Conference on Scale-Space Theories in Computer Vision*. 1999.

[22] M. Bertalmıo, L. Cheng, S. Osher and G. Sapiro. "Variational Problems and Partial Differential Equations on Implicit Surfaces", *Journal of Computational Physics*. 2001

[23] T. Nguyen, J. Cai, J. Zhang, J. Zheng. "Robust Interactive Image Segmentation Using Convex Active Contours", *IEEE Transactions on Image Processing*. 2012

[24] J. Doyle, B. Francis, and A. Tannenbaum. "Feedback Control Theory", *Courier Corporation*. 2013.

[25] R. Sandhu, S. Dambreville, A. Yezzi, A. Tannenbaum. "A Nonrigid Kernel-Based Framework for 2D3D Pose Estimation and 2D image segmentation", *IEEE TPAMI*. 2010.

[26] S. Kichenassamy, A. Kumar, P. Olver, A.Tannenbaum, A. Yezzi. "Conformal Curvature Flows: From Phase Transitions to Active Vision", *Archive for Rational Mechanics and Analysis*. 1996.

[27] M. Do Carmo. "Differential Geometry of Curves and Surfaces: Revised and Updated Second Edition", *Courier Dover Publications*. 2016.

[28] R. Sandhu, T Georgiou, E Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum. "Graph Curvature for Differentiating Cancer Networks", *Scientific reports*. 2015.

[29] B. Bamieh, F. Paganini, M Dahleh. "Distributed Control of Spatially Invariant Systems" *IEEE TAC*. 2002.