Three-Filters-to-Normal+: Revisiting Discontinuity Discrimination in Depth-to-Normal Translation

Jingwei Yang, *Graduate Student Member, IEEE*, Bohuan Xue, *Graduate Student Member, IEEE*, Yi Feng, *Graduate Student Member, IEEE*, Deming Wang, *Member, IEEE*, Rui Fan, *Senior Member, IEEE*, Qijun Chen, *Senior Member, IEEE*

***Abstract*—This article introduces three-filters-to-normal+ (3F2N+), an extension of our previous work three-filters-to-normal (3F2N), with a specific focus on incorporating discontinuity discrimination capability into surface normal estimators (SNEs). 3F2N+ achieves this capability by utilizing a novel discontinuity discrimination module (DDM), which combines depth curvature minimization and correlation coefficient maximization through conditional random fields (CRFs). To evaluate the robustness of SNEs on noisy data, we create a large-scale synthetic surface normal (SSN) dataset containing 20 scenarios (ten indoor scenarios and ten outdoor scenarios with and without random Gaussian noise added to depth images). Extensive experiments demonstrate that 3F2N+ achieves greater performance than all other geometry-based surface normal estimators, with average angular errors of 7.85°, 8.95°, 9.25°, and 11.98° on the clean-indoor, clean-outdoor, noisy-indoor, and noisy-outdoor datasets, respectively. We conduct three additional experiments to demonstrate the effectiveness of incorporating our proposed 3F2N+ into downstream robot perception tasks, including freespace detection, 6D object pose estimation, and point cloud completion. Our source code and datasets are publicly available at https://mias.group/3F2Nplus.**

***Note to Practitioners*—The primary motivation behind this work arises from the need to develop a high-performing surface normal estimator for practical robotics and computer vision applications. While geometry-based surface normal estimators have been widely used in these domains, the existing solutions focus merely on discontinuity discrimination. To tackle this problem, this article introduces a plug-and-play module that leverages both depth curvature and correlation coefficient to quantify discontinuity levels, thereby optimizing surface normal estimation, particularly near or on discontinuous regions. Moreover, this article also introduces a large-scale public dataset with random noise added to depth images, providing a more realistic and robust platform for algorithm evaluation within this research community. Extensive experimental results demonstrate that our method outperforms other state-of-the-art algorithms.**

This research was supported by the National Key R&D Program of China under Grant 2020AAA0108100, the National Natural Science Foundation of China under Grant 62233013, the Science and Technology Commission of Shanghai Municipal under Grant 22511104500, and the Fundamental Research Funds for the Central Universities.

Jingwei Yang, Yi Feng, Deming Wang, Rui Fan, and Qijun Chen are with the College of Electronics & Information Engineering, Shanghai Research Institute for Intelligent Autonomous Systems, the State Key Laboratory of Intelligent Autonomous Systems, and Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 201804, China. (e-mails: {jw_yang, fengyi0109, wangdeming, rfan, qjchen}@tongji.edu.cn)

Bohuan Xue is with the Department of Computer Science & Engineering, the Hong Kong University of Science and Technology, Hong Kong SAR, China. (e-mail: bxueaa@connect.ust.hk)

***Index Terms*—discontinuity discrimination, surface normal, depth curvature, correlation coefficient, conditional random fields, robot perception.**

I. Introduction

As an informative visual feature representing planar characteristics, surface normal has been prevalently utilized in numerous computer vision and robotics applications, such as collision-free space detection [1]–[4], point cloud completion [5], and 6D object pose estimation [6]. Existing surface normal estimators (SNEs) have predominantly employed optimization techniques [7]–[10], *e.g.*, singular value decomposition (SVD) [11] and principal component analysis (PCA) [12] to process unstructured range sensor data. Such algorithms are, nevertheless, computationally intensive in nature [13]. Therefore, there has been a shift in focus towards structured range sensor data, specifically depth images. This category of algorithms is typically referred to as "*depth-to-normal translator*". Our previously published works, *e.g.*, three-filters-to-normal (3F2N) [13], depth-to-normal translator (D2NT) [14], and spatial discontinuity-aware SNE (SDA-SNE) [15], leverage basic image filters to perform depth-to-normal translation. Such approaches stand as the state-of-the-art (SoTA) in the domain of surface normal estimation.

The performance of an SNE is profoundly impacted by the quality of the raw sensor data (point clouds or depth maps) it processes [14]. Such sensor data often exhibit substantial variations, especially in regions with discontinuities [14]. Therefore, the primary objective of this study is to introduce innovative approaches to optimize surface normals in these discontinuous regions. Additionally, it is important to note that for real-world datasets, such as NYU2 [16] and KITTI [17], surface normal ground truth is often unavailable [14]. Their so-called "ground truth" is derived by estimating surface normals directly from noisy depth images or 3D point clouds using a traditional SNE. While we have provided a public dataset alongside 3F2N in [13], this dataset is devoid of any noise and cannot accurately reflect the performance of SNEs in realistic scenarios.

To address the existing challenges mentioned above, our primary focus in this article is to enhance the discontinuity discrimination capability of our previous work 3F2N. The main contributions of this work can be summarized as follows:

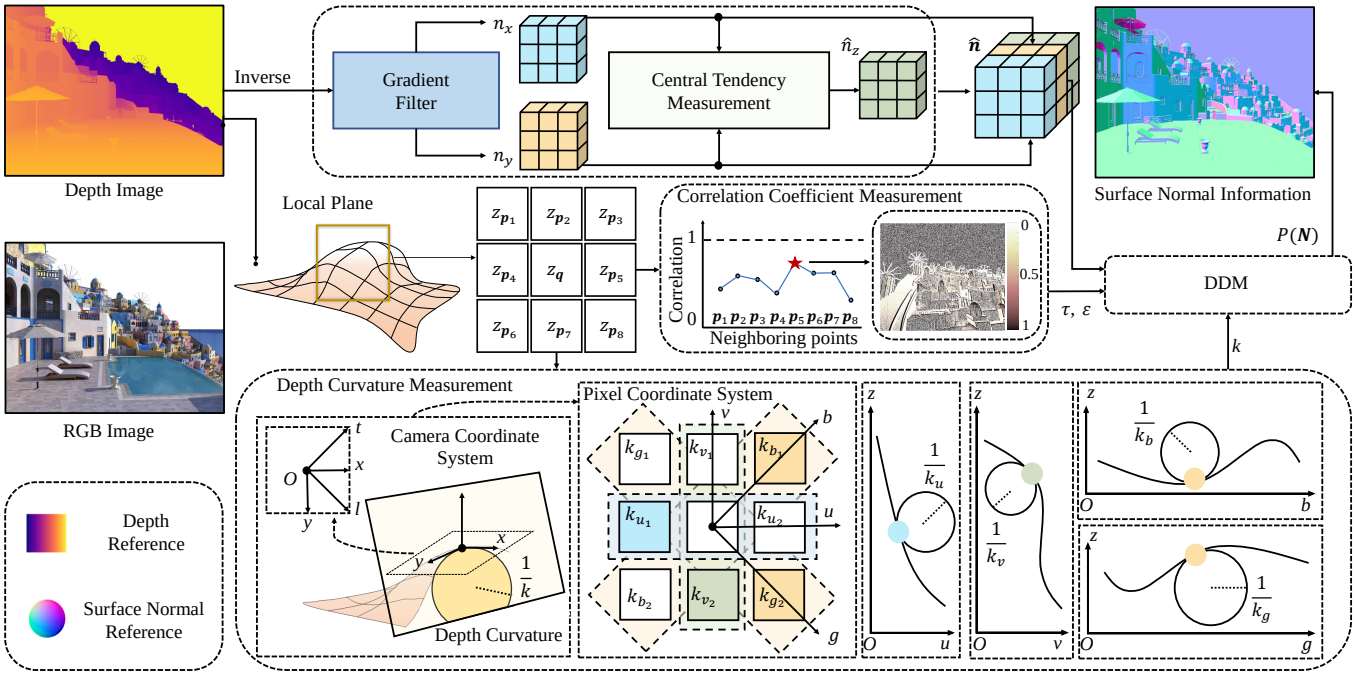

Fig. 1. The workflow of our proposed 3F2N+ SNE.

- An exploration into the feasibility of utilizing depth curvature and correlation coefficient for discontinuity extent quantification.
- A novel discontinuity discrimination module (DDM), which incorporates the measured discontinuity extent into the feature function of conditional random field (CRF), effectively distinguishing discontinuities and thereby optimizing surface normal estimation. The 3F2N integrated with DDM is denoted as 3F2N+ in this article.
- A large-scale synthetic surface normal (SSN) dataset consisting of 20K depth images (both clean and noisy) and their corresponding surface normal ground truth. This dataset is designed to comprehensively evaluate SNE performance in both indoor and outdoor scenarios, with an even split of ten scenarios each.
- A series of additional experiments underscore the benefits of incorporating our proposed 3F2N+ SNE into three downstream computer vision and robotics tasks: (1) data-fusion freespace detection, (2) 6D object pose estimation, and (3) point cloud completion.

The remainder of this article is structured as follows: Sect. II provides a comprehensive review of the SoTA SNEs. Sect. III introduces our proposed 3F2N+ SNE. Extensive experiments demonstrating the robustness of our approach are detailed in Sect. IV. Sect. V delves into a range of computer vision and robotics applications assisted with 3F2N+ SNE. Finally, we summarize our work in Sect. VI.

II. Literature Review

In this section, we provide a comprehensive overview of the SoTA surface normal estimators. A 3D point is denoted as $\boldsymbol{q} = (x; y; z)$, and its corresponding surface normal is denoted as $\boldsymbol{n} = (n_x; n_y; n_z)$. A set storing the neighboring points of $\boldsymbol{q}$ is denoted as $\boldsymbol{P} = (\boldsymbol{p}_1, ..., \boldsymbol{p}_k)$. An augmented set of $\boldsymbol{P}$ is denoted as $\boldsymbol{P}^+ = (\boldsymbol{q}, \boldsymbol{P})$.

A. Point Cloud-to-Normal SNEs

In this category of SNEs[1], surface normals are computed from 3D point clouds by determining the normal vectors of locally interpolated planar surfaces. One common approach formulates this process as the following optimization problem:

$$\arg\min_{\boldsymbol{n}} \; J(\boldsymbol{q}, \boldsymbol{P}, \boldsymbol{n}), \tag{1}$$

where $J(\cdot)$ denotes the cost function, which takes various criteria into account depending on the specific method being used. For instance, it may quantify the plane fitting error in *PlaneSVD* [18] and *PlanePCA* [19], the angle difference in *VectorSVD* [7], or the plane fitting error in the spherical coordinate system in *FALS* [9]. Another solution to this problem is to average the surface normals from neighboring triangles, which can be formulated as follows:

$$\boldsymbol{n} = \frac{1}{k}\sum_{i=1}^{k} \omega_i \frac{[\boldsymbol{p}_i - \boldsymbol{q}]_\times (\boldsymbol{p}_{i+1} - \boldsymbol{q})}{\left\| [\boldsymbol{p}_i - \boldsymbol{q}]_\times (\boldsymbol{p}_{i+1} - \boldsymbol{q}) \right\|_2}, \tag{2}$$

where ω_i denotes the weight associated with the given triangle. *AreaWeighted* [20] determines ω_i based on the magnitude of each triangle, while *AngleWeighted* [20] determines ω_i based on the angle between vectors $\boldsymbol{p}_i - \boldsymbol{q}$ and $\boldsymbol{p}_{i+1} - \boldsymbol{q}$. *SRI* [9] computes surface normals using the following expression:

$$\boldsymbol{n} = (\hat{\boldsymbol{z}}; \hat{\boldsymbol{x}}; \hat{\boldsymbol{y}})\boldsymbol{R}(1; \frac{\partial r}{r\cos\phi\partial\theta}; \frac{\partial r}{r\partial\phi}), \tag{3}$$

[1]Point cloud-to-normal approaches can also be employed to estimate surface normals from depth images when the camera intrinsic matrix is known.

where $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ denote the unit vectors along the respective coordinate directions, respectively, and r, θ, and ϕ denote the range, azimuth, and elevation components in the spherical coordinate system, respectively, and $\boldsymbol{R}$ is a rotation matrix.

B. Depth-to-Normal SNEs

LINE-MOD [21] is a pioneering depth-to-normal SNE. It starts by computing the optimal depth gradient and then forms a 3D plane using three triangle vertices. Finally, the surface normal can be obtained as the tangential vector of the 3D plane. In our previous works 3F2N [13], SNE-RoadSeg [1], and SNE-RoadSeg+ [3], we utilize basic image gradient filters to compute the components n_x and n_y of surface normals:

$$n_x = f_x \frac{\partial 1/z}{\partial u} = f_x g_u, \quad n_y = f_y \frac{\partial 1/z}{\partial v} = f_y g_v, \tag{4}$$

where f_x and f_y denote the camera focal lengths (in pixels) in the x and y directions, respectively. SNE-RoadSeg computes n_z using the following expression [22]:

$$n_z = \cos(\arctan \frac{\sum_{i=1}^{k} \bar{n}_{x,i} \cos\phi + \sum_{i=1}^{k} \bar{n}_{y,i} \sin\phi}{\sum_{i=1}^{k} \bar{n}_{z,i}}), \tag{5}$$

where $\bar{\boldsymbol{n}}_i = \frac{\boldsymbol{n}_i}{||\boldsymbol{n}_i||_2} = (\bar{n}_{x,i}; \bar{n}_{y,i}; \bar{n}_{z,i})$ and $\phi = \arctan(\frac{f_y g_v}{f_x g_u})$. 3F2N computes n_z using the following expression:

$$n_z = -\Phi \left\{ \frac{\Delta x_i n_x + \Delta y_i n_y}{\Delta z_i} \right\}, \tag{6}$$

where $\Phi\{\cdot\}$ denotes a central tendency measurement approach and $(\Delta x_i; \Delta y_i; \Delta z_i) = \boldsymbol{p}_i - \boldsymbol{q}$. Our recent works D2NT [14] and SDA-SNE [15] compute surface normals in an end-to-end manner as follows:

$$\boldsymbol{n}_i = (f_x \frac{\partial z}{\partial u}; f_y \frac{\partial z}{\partial v}; -z - \frac{(u-u_0)\partial z}{\partial u} - \frac{(v-v_0)\partial z}{\partial v}), \tag{7}$$

where $\boldsymbol{p}_0 = (u_0; v_0)$ is the image principal point.

III. Methodology

The framework is schematically illustrated in Fig. 1. A 3×3 local region contains the depth $z_{\boldsymbol{q}}$ of the target point $\boldsymbol{q}$, along with depths $z_{\boldsymbol{p}_i}$ of the eight neighboring points $\boldsymbol{p}_i$ ($i = [1, 8] \bigcap \mathbb{Z}^+$). We introduce two methods to enhance the accuracy of surface normal estimation near/on discontinuities. One is based on the depth curvature minimization (see Sect. III-A), while the other is based on the correlation coefficient maximization (see Sect. III-B). Finally, we present an optimization strategy (see Sect. III-C) that integrates these two modules and can be embedded into all SNEs.

A. Depth Curvature-Based Discontinuity Discrimination

In this subsection, we aim to quantify local surface smoothness using the depth curvature k of neighboring points. A lower depth curvature corresponds to a smoother surface [23]. To comprehensively discriminate discontinuities, we calculate four types of curvatures: mean, maximum, normal, and Gaussian. Mean and maximum depth curvatures are typically obtained by computing the first- and second-order partial derivatives of depth. Therefore, our initial step involves computing the first-order partial derivatives of depth in both the x and y directions within the camera coordinate system as follows:

$$\begin{cases} c_x = \dfrac{\partial z}{\partial x} = \dfrac{f_x}{(u-u_0) + z e_u}, \\ c_y = \dfrac{\partial z}{\partial y} = \dfrac{f_y}{(v-v_0) + z e_v}, \end{cases} \tag{8}$$

where $e_u = \frac{\partial u}{\partial z}$ and $e_v = \frac{\partial v}{\partial z}$ represent the inverse partial derivatives of depth in the u and v directions within the pixel coordinate system, respectively.

In addition to computing partial derivatives in the horizontal and vertical directions (*i.e.*, the x and y directions) in the camera coordinate system, we extend the computations to include first-order partial derivatives in the diagonal directions (*i.e.*, the l and t directions as depicted in Fig. 1). These derivatives are denoted as c_l and c_t, respectively, and are calculated as follows:

$$\begin{cases} c_l = \dfrac{\partial z}{\partial l} = \dfrac{\sqrt{2}}{2}(\dfrac{f_x}{(u-u_0) + z e_u} + \dfrac{f_y}{(v-v_0) + z e_v}), \\ c_t = \dfrac{\partial z}{\partial t} = \dfrac{\sqrt{2}}{2}(\dfrac{f_x}{(u-u_0) + z e_u} - \dfrac{f_y}{(v-v_0) + z e_v}). \end{cases} \tag{9}$$

where the subscripts l and t represent the leading diagonal and trailing diagonal, respectively.

Within the camera coordinate system, the second-order partial derivatives in the horizontal, vertical, and diagonal directions are subsequently achieved by the second-order partial derivatives of depth in the u and v directions as follows:

$$\begin{cases} c_{xx} = \dfrac{\partial^2 z}{\partial x^2} = -\dfrac{f_x^2 \alpha_0}{\alpha_1^3}, \quad c_{yy} = \dfrac{\partial^2 z}{\partial y^2} = -\dfrac{f_y^2 \beta_0}{\beta_1^3}, \\ c_{ll} = \dfrac{\partial^2 z}{\partial l^2} = \dfrac{\sqrt{2}}{2}(c_{xx} + c_{yy}) = -\dfrac{\sqrt{2}}{2}(\dfrac{f_x \alpha_0}{\alpha_1^3} + \dfrac{f_y \beta_0}{\beta_1^3}), \\ c_{tt} = \dfrac{\partial^2 z}{\partial t^2} = \dfrac{\sqrt{2}}{2}(c_{xx} - c_{yy}) = \dfrac{\sqrt{2}}{2}(\dfrac{f_y \beta_0}{\beta_1^3} - \dfrac{f_x \alpha_0}{\alpha_1^3}), \end{cases} \tag{10}$$

where

$$\begin{cases} \alpha_0 = e_u(2 - z e_u^2 e_{uu}), \alpha_1 = u - u_0 + z e_u, e_{uu} = \dfrac{\partial^2 z}{\partial u^2}, \\ \beta_0 = e_v(2 - z e_v^2 e_{vv}), \beta_1 = v - v_0 + z e_v, e_{vv} = \dfrac{\partial^2 z}{\partial v^2}. \end{cases} \tag{11}$$

As shown in Fig. 1, the l and t directions in the camera coordinate system respectively correspond to the g and b directions in the pixel coordinate system. We define the depth curvature in each direction in the 3×3 local region as k_u, k_v, k_g, and k_b, which can be computed as follows:

$$\begin{cases} k_u = \dfrac{|c_{xx}|}{(1 + {c_x}^2)^{\frac{3}{2}}} = |\dfrac{f_x^2 \alpha_0}{(\alpha_1^2 + f_x^2)^{\frac{3}{2}}}|, \\ k_v = \dfrac{|c_{yy}|}{(1 + {c_y}^2)^{\frac{3}{2}}} = |\dfrac{f_y^2 \beta_0}{(\beta_1^2 + f_y^2)^{\frac{3}{2}}}|, \\ k_g = \dfrac{|c_{ll}|}{(1 + {c_l}^2)^{\frac{3}{2}}} = 2|\dfrac{f_x \alpha_0 \beta_1^3 + f_y \beta_0 \alpha_1^3}{[2\alpha_1^2 \beta_1^2 + (f_x \beta_1 + f_y \alpha_1)^2]^{\frac{3}{2}}}|, \\ k_b = \dfrac{|c_{tt}|}{(1 + {c_t}^2)^{\frac{3}{2}}} = 2|\dfrac{f_y \beta_0 \alpha_1^3 - f_x \alpha_0 \beta_1^3}{[2\alpha_1^2 \beta_1^2 + (f_x \beta_1 + f_y \alpha_1)^2]^{\frac{3}{2}}}|. \end{cases} \tag{12}$$

Among the eight neighboring points in the pixel coordinate system, we assign the curvatures of points in the horizontal and vertical directions as k_{u_1}, k_{u_2}, k_{v_1}, and k_{v_2}, the curvatures of the two points in the leading diagonal direction as k_{g_1} and k_{g_2}, and the curvatures of the two points in the trailing diagonal direction as k_{b_1} and k_{b_2}. The maximum curvature $k_{\max}$ at point $\boldsymbol{q}$ on the surface can therefore be obtained as follows:

$$k_{\max} = \max_{i=1,2}(k_{u_i}, k_{v_i}, k_{g_i}, k_{b_i}). \tag{13}$$

Similarly, the mean curvature k_{mean} at point $\boldsymbol{q}$ can be computed as follows:

$$k_{\text{mean}} = \sum_{i=1}^{2}(k_{u_i} + k_{v_i} + k_{g_i} + k_{b_i})/8. \tag{14}$$

In contrast to the maximum and mean depth curvatures, which concentrate on measuring the curvature of a surface at a given point, normal and Gaussian depth curvatures are used to depict the surface smoothness in a specific direction through the computations of the first and second fundamental forms:

$$k_{\text{normal}} = \frac{Le_v^2 + 2Me_ue_v + Ne_u^2}{Ee_v^2 + 2Fe_ue_v + Ge_u^2}, \quad k_{\text{gauss}} = \frac{LN - M^2}{EG - F^2}, \tag{15}$$

where E, F, and G are the coefficients of the first fundamental form. L, M, and N are the coefficients of the second fundamental form. These terms can be computed as follows:

$$\begin{cases} L = c_{xx}c_{yy} = \dfrac{f_x^2 f_y^2 \alpha_0 \beta_0}{\alpha_1^3 \beta_1^3}, \\ M = c_{xx}c_{ll} = \dfrac{\sqrt{2}}{2}(\dfrac{f_x^3 \alpha_0^2}{\alpha_1^6} + \dfrac{f_x^2 f_y \alpha_0 \beta_0}{\alpha_1^3 \beta_1^3}), \\ N = c_{tt}c_{yy} = \dfrac{\sqrt{2}}{2}(\dfrac{f_x f_y^2 \alpha_0 \beta_0}{\alpha_1^3 \beta_1^3} - \dfrac{f_y^3 \beta_0^2}{\beta_1^6}), \\ E = c_x^2 = \dfrac{f_x^2}{\alpha_1^2}, G = c_y^2 = \dfrac{f_y^2}{\beta_1^2}, F = c_x c_y = \dfrac{f_x f_y}{\alpha_1 \beta_1}. \end{cases} \tag{16}$$

The four depth curvatures are employed to quantify the local plane's smoothness. A higher depth curvature indicates a steeper surface within the point's neighborhood, implying a higher probability of discontinuities. Consequently, the optimum surface normal $\boldsymbol{n_q}$ can be determined by considering the depth curvature of the target point $\boldsymbol{q}$ along with its neighboring points. The surface normal estimation for the target point is represented by the surface normal of the neighboring point $\boldsymbol{r}$ with the smallest depth curvature, as follows:

$$\boldsymbol{r} = \arg\min_{\boldsymbol{r} \in \boldsymbol{P}^+} k_{\boldsymbol{r}}, \quad \boldsymbol{n_q} = \boldsymbol{n_r}, \tag{17}$$

where we have the flexibility to choose any one of the curvatures $k_{\max}$, k_{mean}, k_{normal}, and k_{gauss} to represent the depth curvature $k_{\boldsymbol{r}}$. This article demonstrates that the surface normal estimation for the target point is well-suited to the plane in which it is located.

B. Correlation Coefficient-Based Discontinuity Discrimination

To comprehensively identify discontinuities, we not only utilize depth curvature but also incorporate Kendall's [24] and Pearson's correlation coefficients [25]. We perform a mathematical analysis of various combinations of depth values for both the target point and its neighboring points. In accordance with the provided depth map, we define point pairs $(z_{\boldsymbol{q}}, z_{\boldsymbol{p}_i})$ for the computation of Kendall's correlation coefficient within a local 3×3 region. Subsequently, we evaluate the consistency of these pairs. Following the calculation principles of Kendall's correlation coefficient, consistent combinations are those in which the depth of the target point is less than the depth of the neighboring points, resulting in a total of μ_c group pairs. Inconsistent combinations, on the other hand, occur when the depth of the target point exceeds that of the neighboring points, resulting in a total of μ_d combinations. It is important to note that if the target point and neighboring points within a point pair share identical depth values, neither consistent nor inconsistent combinations are considered.

We define Kendall's correlation coefficient for discriminating discontinuous regions as τ, which is determined by the ratio of consistent to inconsistent pairs. This coefficient ranges from 0 to 1, with larger values indicating a higher probability that the points belong to the same continuous plane. Consequently, as the coefficient increases, the surface normal of the target point becomes more aligned with the surface normal of neighboring points. The final surface normal $\boldsymbol{n_q}$ for the target point is determined by selecting the neighboring point $\boldsymbol{r}$ with the maximum Kendall's correlation coefficient $\tau_{\boldsymbol{r}}$ as follows:

$$\begin{cases} \tau_{\boldsymbol{r}} = |\dfrac{\mu_c - \mu_d}{\mu_c + \mu_d}|, \boldsymbol{r} \in \boldsymbol{P}^+, \\ \boldsymbol{r} = \arg\max\limits_{\boldsymbol{r} \in \boldsymbol{P}^+} \tau_{\boldsymbol{r}}, \quad \boldsymbol{n_q} = \boldsymbol{n_r}. \end{cases} \tag{18}$$

To calculate Pearson's correlation coefficient $\varepsilon_{\boldsymbol{r}}$ for point $\boldsymbol{r}$, we begin by constructing a new sequence composed of the depths of neighboring points, denoted as $\boldsymbol{\psi} = (z_{\boldsymbol{p}_1}, ..., z_{\boldsymbol{p}_8})$. We then create a sequence $\boldsymbol{\psi}^+ = (z_{\boldsymbol{q}}, z_{\boldsymbol{p}_1}, ..., z_{\boldsymbol{p}_8})$. The Pearson's correlation coefficient can be calculated as follows:

$$\begin{cases} \varepsilon_{\boldsymbol{r}} = \dfrac{|(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi}^+ - \bar{\boldsymbol{\psi}}^+)|}{\sqrt{||\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}||^2 ||\boldsymbol{\psi}^+ - \bar{\boldsymbol{\psi}}^+||^2}}, \boldsymbol{r} \in \boldsymbol{P}^+, \\ \boldsymbol{r} = \arg\max\limits_{\boldsymbol{r} \in \boldsymbol{P}^+} \varepsilon_{\boldsymbol{r}}, \quad \boldsymbol{n_q} = \boldsymbol{n_r}, \end{cases} \tag{19}$$

where the coefficient $\varepsilon_{\boldsymbol{r}}$ ranges from 0 to 1, and $\boldsymbol{n_q}$ represents the optimized surface normal via discontinuity discrimination.

C. Discontinuity Discrimination Module

We introduce a CRF-based surface normal optimization method that combines the two discontinuity discrimination strategies, respectively discussed in Sect. III-A and Sect. III-B, to further minimize the surface normal estimation error in discontinuous regions.

For a given depth image, we first define an undirected graph $\boldsymbol{G} = (\boldsymbol{Q}, \boldsymbol{E})$, where the nodes $\boldsymbol{Q}$ denote pixels in the depth image, while the edges $\boldsymbol{E}$ denote the correlations between

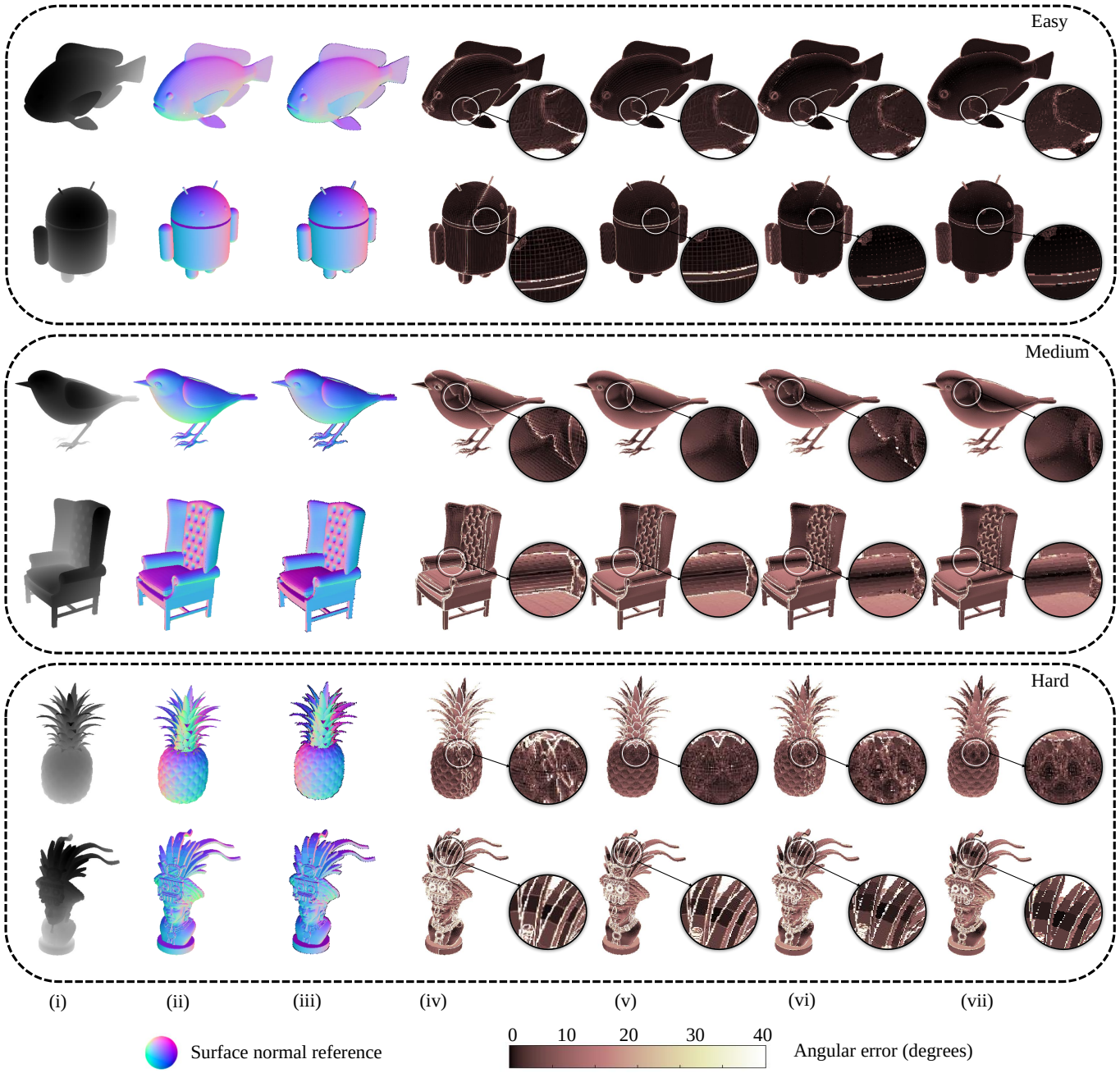

Fig. 2. Examples of surface normal estimation results on the 3F2N-Easy, 3F2N-Medium and 3F2N-Hard datasets. (i) and (ii) show depth and surface normal ground truth, respectively. (iii) shows the surface normal estimation results achieved by 3F2N+. (iv)-(vii) show angular error maps obtained by AngleWeighted, 3F2N, AngleWeighted+, and 3F2N+, respectively.

adjacent pixels. We then define $P(\boldsymbol{N})$ as a conditional probability distribution that characterizes the surface normal $\boldsymbol{n_q}$ within the surface normal field $\boldsymbol{N}$ as follows:

$$P(\boldsymbol{N}) = \frac{1}{\Theta(\boldsymbol{N})} \exp\left(\sum_{\boldsymbol{q}\in\boldsymbol{Q}} \sigma f(\boldsymbol{q}) + \sum_{(\boldsymbol{p}_i,\boldsymbol{q})\in\boldsymbol{E}} \lambda_i g(\boldsymbol{p}_i,\boldsymbol{q})\right), \tag{20}$$

where $\Theta(\boldsymbol{N})$ is the partition function that normalizes the conditional probability distribution, σ and λ_i represent weights of $f(\cdot)$ and $g(\cdot)$, respectively.

$$f(\boldsymbol{q}) = \|\boldsymbol{n_q} - \hat{\boldsymbol{n}}_{\boldsymbol{q}}\| \tag{21}$$

is the node feature function used to compute the cost of the surface normal of the point $\boldsymbol{q}$, thereby finding the best estimation of the surface normal $\hat{\boldsymbol{n}}_{\boldsymbol{q}}$.

$$g(\boldsymbol{p}_i,\boldsymbol{q}) = k_{\boldsymbol{q}} \sum_i (1 - \tau_{\boldsymbol{p}_i} \vee \varepsilon_{\boldsymbol{p}_i}) \|\boldsymbol{n}_{\boldsymbol{p}_i} - \hat{\boldsymbol{n}}_{\boldsymbol{q}}\| \tag{22}$$

denotes the edge feature function, which is used to penalize the surface normal discontinuity among pairs of neighboring

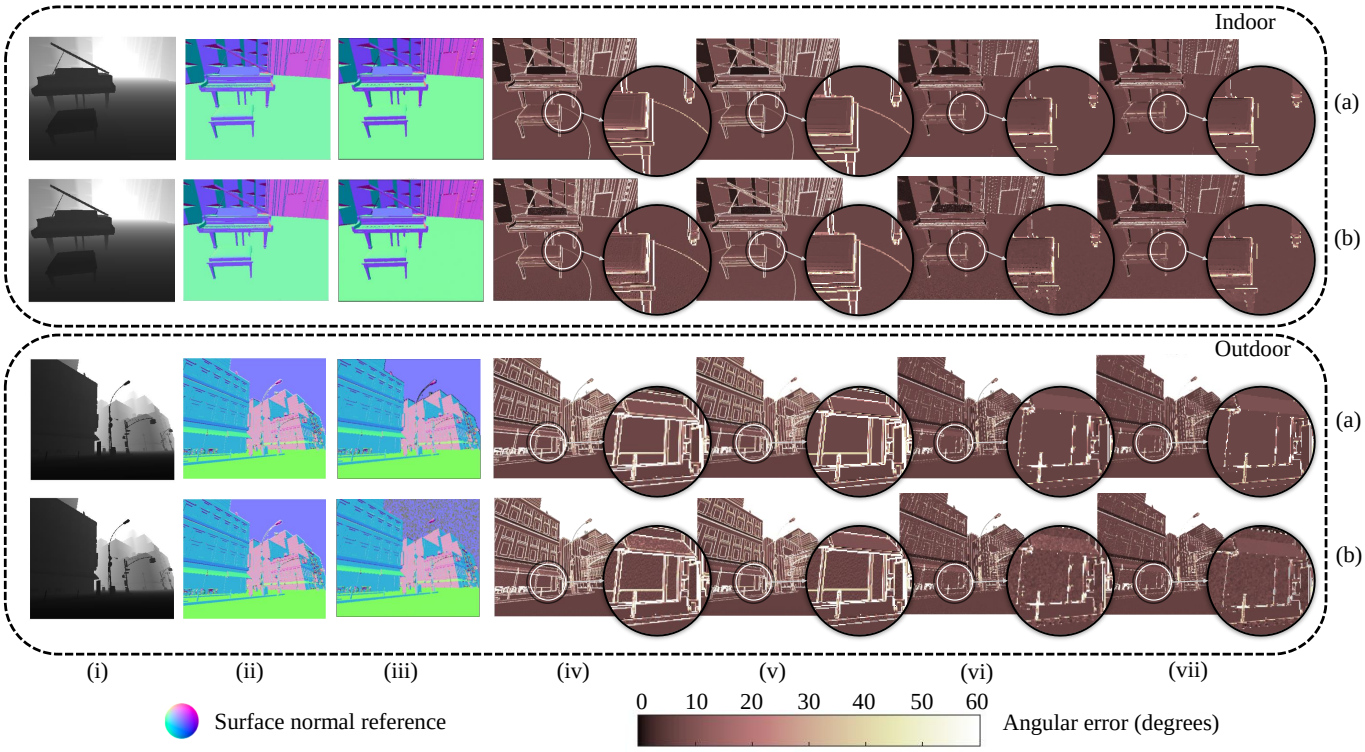

Fig. 3. Examples of surface normal estimation results on the SSN dataset. (i) and (ii) show depth and surface normal ground truth, respectively. (iii) shows the surface normal estimation results achieved by 3F2N+. (iv)-(vii) show angular error maps obtained by AngleWeighted, 3F2N, AngleWeighted+, and 3F2N+, respectively. (a) and (b) show the results with respect to clean and noisy inputs, respectively.

pixels, where $k_{\boldsymbol{q}}$ is the depth curvature obtained in Sect. III-A. (22) illustrates that when neighboring points tend to align with a plane, it is more probable for the distribution of surface normals to exhibit similarity. As a result, this leads to optimized surface normals that are more aligned with the underlying surfaces.

IV. Experiments

A. Datasets and Evaluation Metrics

In our previous study [1], we introduced a large-scale, noise-less, synthetic dataset consisting of three subsets: 3F2N-Easy, 3F2N-Medium, and 3F2N-Hard. In this article, we create SSN dataset using 20 3D mesh models, including depth images with and without the addition of random Gaussian noise. Our dataset is divided into two subsets: indoor and outdoor scenarios. For each scenario model, we capture 500 different views using an Intel Core i7-12700H CPU. Z-Buffer rendering [26] is employed to generate depth images with a resolution of 512×424 pixels. All scenario models are represented by meshes composed of triangles, the surface normals of which serve as the ground truth. The camera parameters used in our dataset are identical to those of the Kinect v2 sensor.

The accuracy of SNEs is quantified using the average angular error:

$$e_{\mathrm{A}} = \frac{1}{m}\sum_{k=1}^{m} \cos^{-1}\frac{\langle \boldsymbol{n}_k, \hat{\boldsymbol{n}}_k \rangle}{||\boldsymbol{n}_k||_2||\hat{\boldsymbol{n}}_k||_2}, \tag{23}$$

where m denotes the number of observed points, while $\boldsymbol{n}_k$ and $\hat{\boldsymbol{n}}_k$ denote the ground truth and estimated surface normals, respectively.

B. Performance Evaluation of Depth Curvature-Based Discontinuity Discrimination Approaches

The evaluation results of the four depth curvature-based discontinuity discrimination methods are presented in Table I. As expected, the results obtained using mean and maximum curvatures demonstrate higher accuracy compared to those obtained using normal and Gaussian curvatures. The mean curvature stands out with the most superior performance, making it the most robust choice in our following experiments. Additionally, 3F2N+ using the finite difference (FD) filter, median filter, and mean curvature, achieves an inference speed of 84 FPS.

C. Performance Evaluation of Correlation Coefficient-Based Discontinuity Discrimination Approaches

The quantitative results of two correlation coefficient-based discontinuity discrimination methods combined with 3F2N+ (denoted as 3F2N+ w/ Pearson and 3F2N+ w/ Kendall) are presented in Table II. The e_{A} scores for 3F2N+ w/ Pearson and 3F2N+ w/ Kendall are 11.33% and 14.82% lower than those achieved using 3F2N on the 3F2N datasets, respectively, with 3F2N+ w/ Kendall showing slightly better performance. Additionally, the inference speeds of 3F2N+ w/ Pearson and 3F2N+ w/ Kendall reach 58 FPS and 68 FPS, respectively.

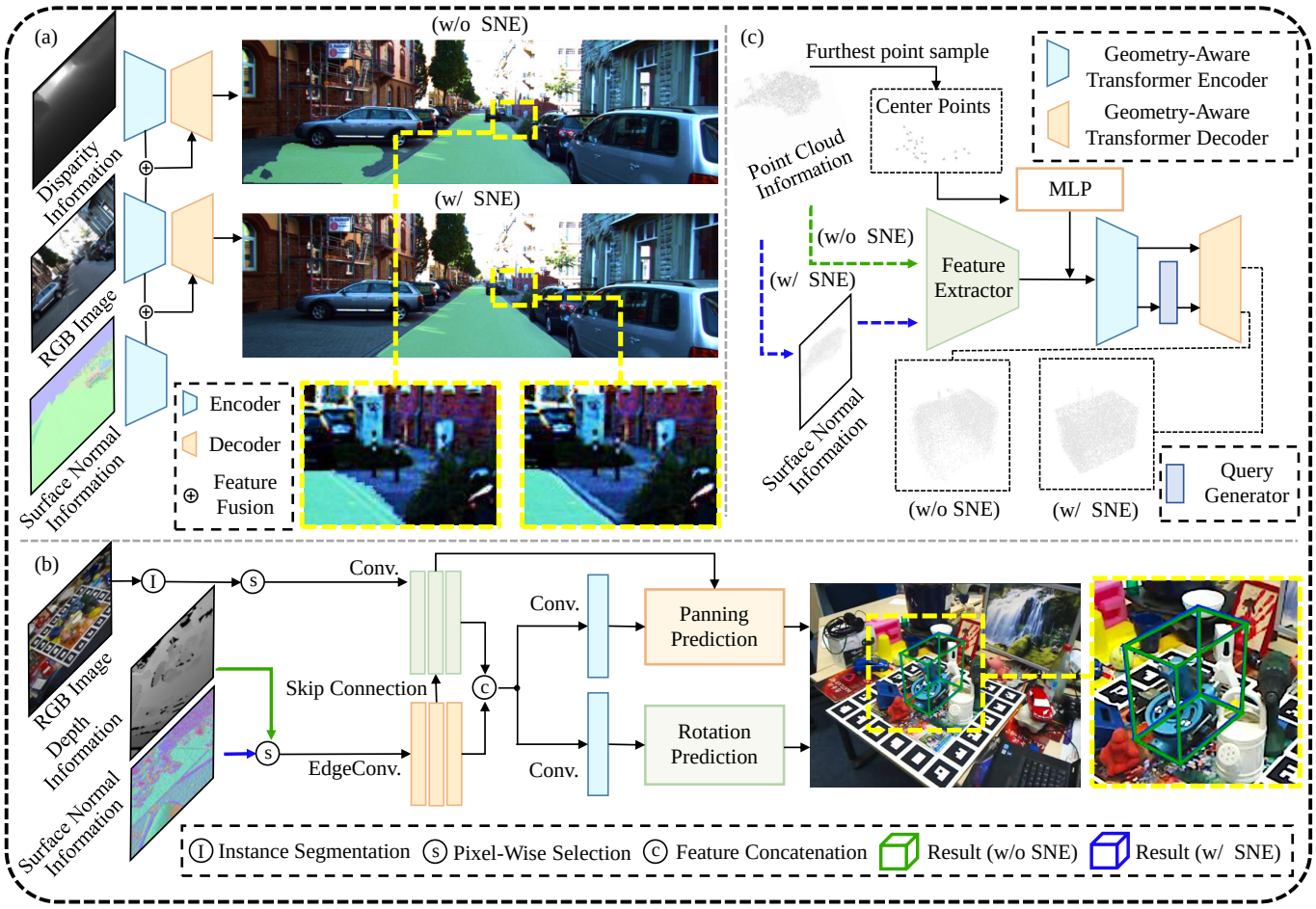

Fig. 4. Robot perception tasks using 3F2N+: (a) data-fusion freespace detection [1]; (b) 6D object pose estimation [6]; (c) point cloud completion [5].

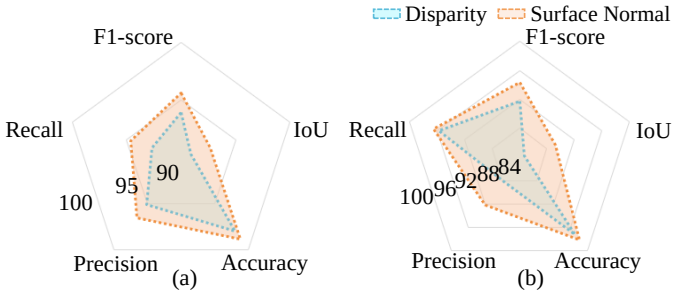

Fig. 5. Performance comparison (%) with respect to different input data for data-fusion freespace detection: (a) MFNet result; (b) FuseNet result.

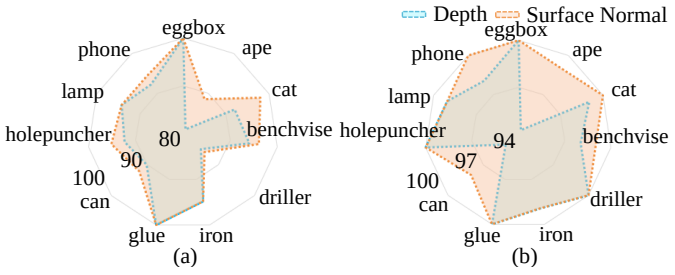

Fig. 6. Performance comparison (%) between two 6D object pose estimation methods: (a) PoseNet; (b) GCN.

D. Performance Evaluation of Discontinuity Discrimination Module

3F2N+ integrates the optimal combination (FD filter, median filter, and mean curvature) obtained in Sect. IV-B along with Kendall's correlation coefficient method mentioned in Sect. IV-C into the DDM. The quantitative results on the 3F2N dataset are presented in Table III. Compared to 3F2N, 3F2N+ achieves a reduction in e_{A} by around 44.17%, 28.50%, and 34.84% on the 3F2N-Easy, 3F2N-Medium, and 3F2N-Hard datasets, respectively. Additionally, we employ AngleWeighted combined with DDM (denoted as AngleWeighted+) to further validate the efficacy of this module. The qualitative results shown in Figs. 2 and 3 demonstrate improved surface normal estimation performance near or on discontinuities with the incorporation of DDM.

E. Performance Evaluation on the 3F2N and SSN Datasets

The average angular error and the runtime comparisons between SoTA SNEs mentioned in Sect. II and 3F2N+ are presented in Table III. The e_{A} scores obtained by 3F2N+ are 0.93°, 4.07°, and 9.98° on the 3F2N-Easy, 3F2N-Medium, and 3F2N-Hard datasets, respectively. Despite achieving SoTA accuracy, 3F2N+ attains an inference speed of 44 FPS.

TABLE I
COMPARISON OF e_A W.R.T. DIFFERENT CENTRAL TENDENCY MEASUREMENTS, GRADIENT FILTERS, AND DEPTH CURVATURES.

Central Tendency Measurement	Gradient Filter	Depth Curvature	e_A (degrees)		
			Easy	Medium	Hard
Median	FD	Normal	1.214	5.321	11.648
		Gaussian	1.216	5.155	11.714
		Maximum	0.914	4.791	9.976
		Mean	**0.905**	**4.776**	**9.959**
	Roberts	Normal	1.343	5.589	11.776
		Gaussian	1.509	5.698	12.197
		Maximum	1.087	5.161	11.186
		Mean	1.036	5.063	10.939
Trimean	FD	Normal	1.228	5.352	11.582
		Gaussian	1.229	5.189	11.342
		Maximum	0.931	4.835	10.013
		Mean	0.922	4.820	9.997
	Roberts	Normal	1.371	5.640	11.873
		Gaussian	1.576	5.808	12.447
		Maximum	1.087	5.161	11.186
		Mean	1.079	5.152	11.155
Trimmean	FD	Normal	1.212	5.305	11.830
		Gaussian	1.518	5.160	11.391
		Maximum	0.939	4.853	10.035
		Mean	0.929	4.839	10.019
	Roberts	Normal	1.704	5.581	12.151
		Gaussian	2.071	5.791	12.763
		Maximum	1.102	5.195	11.245
		Mean	1.095	5.187	11.219

TABLE II
e_A COMPARISONS BETWEEN 3F2N AND 3F2N+ W/ CORRELATION COEFFICIENT-BASED DISCONTINUITY DISCRIMINATION

Method	e_A (degrees)		
	Easy	Medium	Hard
3F2N	1.657	5.685	15.313
3F2N+ w/ Pearson	**1.320**	5.645	13.122
3F2N+ w/ Kendall	1.352	**5.360**	**12.584**

In addition, the quantitative and qualitative results of 3F2N+ on the SSN dataset are shown in Table III and Fig. 3, respectively, and our proposed method obtains the lowest e_A scores, which demonstrates the robustness of 3F2N+ on noisy data. All results indicate that 3F2N+ greatly improves the accuracy of surface normal estimation while still ensuring real-time performance.

V. APPLICATION OF 3F2N+ IN OTHER ROBOT PERCEPTION TASKS

In this section, we incorporate our proposed 3F2N+ into three robot perception tasks [27]: (1) data-fusion freespace detection, (2) 6D object pose estimation, and (3) point cloud completion, to further demonstrate its applicability.

First, we train two data-fusion networks: MFNet [28] and FuseNet [29], on the KITTI Road dataset [30]. For each of these networks, we use two different inputs: RGB+Disparity and RGB+Normal. The results of our experiments are presented in Figs. 4(a) and 5. Notably, when we incorporate surface normal information as the input, we observe significant improvements in performance metrics. Specifically, the intersection over union (IoU) metric shows an increase of up to 6%, and the F1-score demonstrates an increase of up to 3%. These findings align with the conclusions drawn in our prior work [31], providing further empirical evidence to support the effectiveness of the SNE used in freespace detection algorithms.

Figs. 4(b) and 6, as well as Table IV present qualitative and quantitative results of 6D object pose estimation on the LineMOD dataset [32] with inputs of RGB+Depth data and RGB+Normal data. Table IV demonstrates that the accuracy is improved by 1.1% and 1.6%, respectively, and the average 3D distance (ADD) [33] decreases by 0.049 mm and 0.670 mm, respectively, when using PoseNet [34] and graph convolutional network (GCN) [35]. These results strongly suggest that our proposed 3F2N+ can serve as an effective component integrated into 6D object pose estimation networks to enhance their performance.

The qualitative and quantitative results of point cloud completion are presented in Fig. 4(c) and Table V, respectively. When incorporating surface normal information, the F1-score increases by 3% compared to using point cloud information only. Additionally, the Chamfer distance [36] based on the L1-norm (abbreviated as CD-l_1) and L2-norm (abbreviated as CD-l_2) decrease by 0.40 mm and 0.05 mm, respectively. These results indicate that our proposed 3F2N+ has a stronger ability to recover details in the point cloud completion task.

VI. CONCLUSION

In this study, we introduced 3F2N+, an extension of 3F2N, enhanced with novel discontinuity discrimination techniques. These techniques leverage depth curvature minimization and correlation coefficient maximization to effectively address surface normal estimation challenges near or on discontinuities. To evaluate the robustness of SNEs on noisy data, we created a large-scale dataset, referred to as SSN, which contains both clean and noisy depth images. Through extensive experiments on the 3F2N and SSN datasets, we demonstrated the superior performance of 3F2N+ compared to existing SNEs. In addition, we explored the versatility of 3F2N+ in data-fusion freespace detection, 6D object pose estimation, and point cloud completion tasks. We believe DDM can serve as a universal solution for enhancing SNEs. Our commitment to ongoing research and development in this field remains unwavering.

REFERENCES

[1] R. Fan *et al.*, "SNE-Roadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 340–356.
[2] J. Li *et al.*, "RoadFormer: Duplex Transformer for RGB-normal semantic road scene parsing," *arXiv preprint arXiv:2309.10356*, 2023.

TABLE III
e_A AND THE RUNTIME COMPARISONS AMONG SOTA SNES ON THE 3F2N AND SSN DATASETS.

Method	Runtime (ms)	e_A (degrees)						
		3F2N dataset			SSN dataset			
		Easy	Medium	Hard	Clean-Indoor	Noisy-Indoor	Clean-Outdoor	Noisy-Outdoor
PlaneSVD [18]	336.63	2.07	6.07	17.59	11.61	13.05	13.69	14.43
PlanePCA [19]	540.29	2.07	6.07	17.59	12.39	13.05	12.91	14.42
VectorSVD [7]	481.58	2.13	6.27	18.01	12.72	13.61	13.43	16.15
AreaWeighted [20]	933.93	2.20	6.27	17.03	11.94	12.59	12.47	14.09
AngleWeighted [20]	883.17	1.79	5.67	13.26	9.51	10.17	10.37	12.36
FALS [9]	3.51	2.26	6.14	17.34	12.11	12.55	12.47	13.78
SRI [9]	10.41	2.64	6.71	19.61	13.82	14.02	13.73	14.79
LINE-MOD [21]	5.50	6.53	9.94	31.45	20.55	20.58	18.51	18.64
SNE-RoadSeg [1]	6.77	2.04	6.28	16.37	9.10	12.68	11.13	16.14
3F2N (mean) [13]	**3.18**	2.14	6.66	15.30	11.07	16.11	12.15	18.16
3F2N (median) [13]	9.38	1.66	5.69	15.31	11.56	13.08	12.28	15.35
3F2N+ (ours)	22.96	**0.93**	**4.07**	**9.98**	**7.85**	**9.25**	**8.95**	**11.98**

TABLE IV
COMPARISONS WITH RESPECT TO DIFFERENT INPUT DATA FOR 6D OBJECT POSE ESTIMATION.

Method	w/ 3F2N+	Accuracy (%)	ADD (mm)
PoseNet		91.920	1.058
	✓	**93.414**	**1.009**
GCN		98.284	6.067
	✓	**99.398**	**5.397**

TABLE V
F1-SCORE, CD-l_1, AND CD-l_2 COMPARISONS W/ AND W/O 3F2N+ INCORPORATED FOR POINT CLOUD COMPLETION.

w/ normal	F1-score (%)	CD-l_1 (mm)	CD-l_2 (mm)
	0.497	11.622	0.577
✓	**0.511**	**11.221**	**0.530**

[3] H. Wang *et al.*, "SNE-RoadSeg+: Rethinking depth-normal translation and deep supervision for freespace detection," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1140–1145.

[4] H. Wang *et al.*, "Dynamic fusion module evolves drivable area and road anomaly detection: A benchmark and algorithms," *IEEE Transactions on Cybernetics*, vol. 52, no. 10, pp. 10 750–10 760, 2022.

[5] Z. Huang *et al.*, "PF-Net: Point Fractal Network for 3D Point Cloud Completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7662–7670.

[6] P. Liu *et al.*, "BDR6D: Bidirectional Deep Residual Fusion Network for 6D Pose Estimation," *IEEE Transactions on Automation Science and Engineering*, pp. 1–12, 2023, doi: 10.1109/TASE.2023.3248843.

[7] K. Klasing *et al.*, "Comparison of surface normal estimation methods for range sensing applications," in *2009 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2009, pp. 3206–3211.

[8] S. Holzer *et al.*, "Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2012, pp. 2684–2689.

[9] H. Badino *et al.*, "Fast and accurate computation of surface normals from range images," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 3084–3091.

[10] B. Zeisl *et al.*, "Discriminatively trained dense surface normal estimation," in *European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 468–484.

[11] D. Kalman, "A singularly valuable decomposition: the SVD of a matrix," *The College Mathematics Journal*, vol. 27, no. 1, pp. 2–23, 1996.

[12] H. Abdi *et al.*, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[13] R. Fan *et al.*, "Three-filters-to-normal: An accurate and ultrafast surface normal estimator," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5405–5412, 2021.

[14] Y. Feng *et al.*, "D2NT: A high-performing depth-to-normal translator," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 360–12 366.

[15] N. Ming *et al.*, "SDA-SNE: Spatial discontinuity-aware surface normal estimation via multi-directional dynamic programming," in *2022 International Conference on 3D Vision (3DV)*. IEEE, 2022, pp. 486–494.

[16] N. Silberman *et al.*, "Indoor segmentation and support inference from RGBD images," in *European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 746–760.

[17] A. Geiger *et al.*, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354–3361.

[18] K. Jordan *et al.*, "A quantitative evaluation of surface normal estimation in point clouds," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2014, pp. 4220–4226.

[19] K. Klasing *et al.*, "Realtime segmentation of range data using continuous nearest neighbors," in *2009 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2009, pp. 2431–2436.

[20] S. Jin *et al.*, "A comparison of algorithms for vertex normal computation," *The Visual Computer*, vol. 21, no. 1, pp. 71–82, 2005.

[21] S. Hinterstoisser *et al.*, "Gradient response maps for real-time detection of textureless objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 876–888, 2011.

[22] R. Fan *et al.*, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Transactions on Image Processing*, vol. 29, pp. 897–908, 2019.

[23] H. Federer, "Curvature measures," *Transactions of the American Mathematical Society*, vol. 93, no. 3, pp. 418–491, 1959.

[24] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[25] Y. Lian *et al.*, "Research on non-contact multi-person heart rate measurement method for intelligent education," in *2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*. IEEE, 2022, pp. 199–205.

[26] N. Greene *et al.*, "Hierarchical z-buffer visibility," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1993, pp. 231–238.

[27] R. Fan *et al.*, "Autonomous driving perception," 2023.

[28] Q. Ha *et al.*, "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5108–5115.

[29] C. Hazirbas *et al.*, "FuseNet: Incorporating depth into semantic segmentation via fusion-based cnn architecture," in *13th Asian Conference on Computer Vision (ACCV), 2016*. Springer, 2017, pp. 213–228.
[30] M. Menze *et al.*, "Object scene flow for autonomous vehicles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3061–3070.
[31] J. Yang *et al.*, "Semantic segmentation for autonomous driving," in *Autonomous Driving Perception: Fundamentals and Applications*. Springer, 2023, pp. 101–137.
[32] S. Hinterstoisser *et al.*, "Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes," in *2011 International Conference on Computer Vision (ICCV)*. IEEE, 2011, pp. 858–865.
[33] Y. He *et al.*, "FFB6D: A full flow bidirectional fusion network for 6D pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 3003–3013.
[34] A. Kendall *et al.*, "PoseNet: A convolutional network for real-time 6-DoF camera relocalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2938–2946.
[35] T. N. Kipf *et al.*, "Semi-Supervised Classification with Graph Convolutional Networks," in *International Conference on Learning Representations (ICLR)*, 2017.
[36] M. A. Butt *et al.*, "Optimum design of Chamfer distance transforms," *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1477–1484, 1998.

Jingwei Yang (Graduate Student Member, IEEE) received the B.Eng. degree in Automation from the Nanjing University of Science and Technology in 2020. She is currently pursuing a Ph.D. degree with the MIAS Group in the Robotics and Artificial Intelligence Lab at Tongji University. Her research interests include spatial information representation and semantic scene understanding.

Bohuan Xue (Graduate Student Member, IEEE) received his B.Eng. degree in Computer Science and Technology from the College of Mobile Telecommunications at Chongqing University of Posts and Telecommunications in 2018. He is currently pursuing a Ph.D. at the Hong Kong University of Science and Technology. His research interests lie in computer vision and robotics.

Yi Feng (Graduate Student Member, IEEE) received the B.E. degree in automation from Tongji University, Shanghai, China, in 2022. He is currently pursuing his M.Sc. degree with the MIAS Group in the College of Electronics and Information Engineering at Tongji University. His research interests include computer vision and deep learning.

Deming Wang (Member, IEEE) received his B.Eng. degree in Automation and Ph.D. in Control Science and Engineering from Tongji University, in 2017 and 2022, respectively. He is currently working at HUAWEI. His research interests span computer vision, deep learning, and robotics.

Rui Fan (Senior Member, IEEE) received the B.Eng. degree in Automation from the Harbin Institute of Technology in 2015 and the Ph.D. degree (supervisors: Prof. John G. Rarity and Prof. Naim Dahnoun) in Electrical and Electronic Engineering from the University of Bristol in 2018. He worked as a Research Associate (supervisor: Prof. Ming Liu) at the Hong Kong University of Science and Technology from 2018 to 2020 and a Postdoctoral Scholar-Employee (supervisors: Prof. Linda M. Zangwill and Prof. David J. Kriegman) at the University of California San Diego between 2020 and 2021. Rui began his faculty career as a Full Research Professor with the College of Electronics & Information Engineering at Tongji University in 2021, and was then promoted to a Full Professor in the same college, as well as at the Shanghai Research Institute for Intelligent Autonomous Systems in 2022. Rui served as an associate editor of ICRA'23 and IROS'23, and as a senior program committee member of AAAI'23/24. Rui is the general chair of the AVVision community and organized several impactful workshops and special sessions in conjunction with WACV'21, ICIP'21/22/23, ICCV'21, and ECCV'22. Rui was named in the Stanford University List of Top 2% Scientists Worldwide in 2022 and 2023, as well as in the Forbes China List of 100 Outstanding Overseas Returnees in 2023. Rui's research interests include computer vision, deep learning, and robotics.

Qijun Chen (Senior Member, IEEE) received the B.S. degree in automation from Huazhong University of Science and Technology, Wuhan, China, in 1987, the M.S. degree in information and control engineering from Xi'an Jiaotong University, Xi'an, China, in 1990, and the Ph.D. degree in control theory and control engineering from Tongji University, Shanghai, China, in 1999. He is currently a Full Professor in the College of Electronics and Information Engineering, Tongji University, Shanghai, China. His research interests include robotics control, environmental perception, and understanding of mobile robots and bioinspired control.