

Probabilistic Framework for the Characterization of Surfaces and Edges in Range Images, with Application to Edge Detection

Antoine Lejeune, Jacques G. Verly, and Marc Van Droogenbroeck

Abstract—We develop a powerful probabilistic framework for the local characterization of surfaces and edges in range images. We use the geometrical nature of the data to derive an analytic expression for the joint probability density function (pdf) for the random variables used to model the ranges of a set of pixels in a local neighborhood of an image. We decompose this joint pdf by considering independently the cases where two real world points corresponding to two neighboring pixels are locally on the same real world surface or not. In particular, we show that this joint pdf is linked to the Voigt pdf and not to the Gaussian pdf as it is assumed in some applications. We apply our framework to edge detection and develop a locally adaptive algorithm that is based on a probabilistic decision rule. We show in an objective evaluation that this new edge detector performs better than prior art edge detectors. This proves the benefits of the probabilistic characterization of the local neighborhood as a tool to improve applications that involve range images.

Index Terms—Range image, surface, probabilistic framework, edge detection, time-of-flight camera, Kinect

1 INTRODUCTION

Characterizing how real world objects and their surfaces are imaged by a camera is the basis of many image processing and computer vision methods. Range cameras, which are cameras that measure the depth between each pixel of the sensor plane and the corresponding element in the scene, bring pertinent information that can be used by algorithms that interpret images. However, many methods dedicated to range images tend to consider these images in the same way as grayscale or color images [15], and this ignores the physics of the process for acquiring range images.

Because 3D information is crucial for many computer vision tasks, several methods for capturing 3D information have been developed, such as methods of stereo-matching [18] and methods that physically measure the range through the use of active sensors. The use of active sensors allows to achieve a high level of precision [4]. Recent technological advances have permitted the production of commodity range cameras capable of producing range images, at high spatial resolution and real-time frame rate, that contain, at each of their pixels, the range of the corresponding element in the scene. These devices include time-of-flight (ToF) cameras [23] and structured-light (SL) cameras such as the Microsoft Kinect 1 [10] for the SL case.

In this paper, we present a probabilistic framework for surfaces and edges in range images. While machine learning algorithms have recently been shown to deal efficiently with range images [13], [33], a probabilistic framework is complementary and could even be combined with machine learning methods. Also, our framework leads to a better

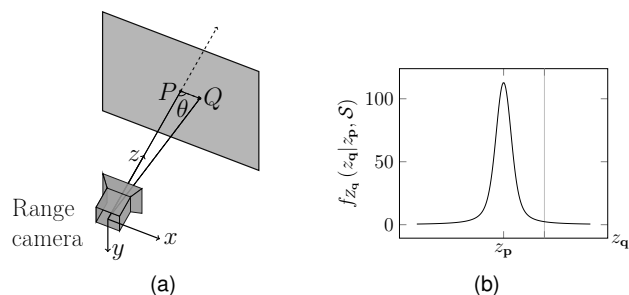


Figure 1. (a) Schematic view of a surface seen by a range camera. P and Q are two points that belong to this surface; they show up at pixels p and q in a corresponding range image. (b) Plot of the probability density function (pdf) of the random variable Z_q that represents the range of q , given the range z_p of p and the assumption that P and Q belong to the same surface. This paper establishes that this pdf is a Voigt pdf and shows how to use this finding for edge detection.

understanding of range images. One of our findings is illustrated in Fig. 1. Also, we show how our framework can be used for the detection of edges, one of the fundamental image descriptors. In this framework, we view the range values of the pixels as realizations of random variables (RVs), and exploit the fact that the geometrical nature of the data produced by range cameras constrains the joint probability density function (pdf) of the range values of two pixels when they correspond to two real world points from the same continuous real world surface, or, inversely, when they straddle a jump edge. We obtain analytic expressions for this joint pdf, which can then be used in applications.

In Section 2, we discuss the prior art for the modeling of range images, with a focus on how edges and surfaces are represented. In Section 3, we determine, using a Bayesian approach, the two joint pdfs of the range values of two pixels under the following two complementary conditions: (1) both pixels correspond to two real world points on the

• A. Lejeune, J. G. Verly, and M. Van Droogenbroeck are with the Dept. of Electr. Engineering and Computer Science, University of Liège, Belgium
E-mail: {Antoine.Lejeune,Jacques.Verly,M.VanDroogenbroeck}@ulg.ac.be

Manuscript received November 7, 2016.

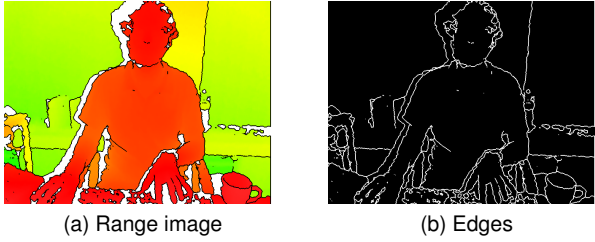


Figure 2. (a) Representation with false colors of a range image obtained with a Microsoft Kinect 1 camera. (b) Corresponding edges obtained by application of an edge detector based on the framework described in this paper.

same surface; (2) both pixels straddle a jump edge. These expressions constitute the core of our framework, and are the main contribution of this paper. In Section 4, we extend our framework to express the joint pdf corresponding to the range values of more than two pixels when we consider that they all correspond to the same real world planar surface. In Section 5, we apply our framework to edge detection; Figure 2 illustrates the edges obtained from a range image. In Section 6, we evaluate the performance of the edge detector based on our new framework and compare it with other edge detectors. Section 7 concludes this paper.

2 STATE-OF-THE-ART

In this paper, we consider the value (called the range or the depth) of each pixel of a range image as being a realization of a random variable (RV), and we determine the probability density function (pdf) of this RV and the joint pdf of the RVs corresponding to the ranges of two neighboring pixels according to a model of the acquisition process and of the real world surfaces. In this section, we review the current state of the art for similar characterizations of the pdf of range values, then the current state of the art for modeling the noise for range cameras, and finally the current state of the art for modeling surfaces and the edges between them.

2.1 Statistics of range images

There have been only a few attempts to characterize the pdf of a RV representing a pixel value in a range image. Huang *et al.* [15] studied the statistics of a database of static forest scenes captured by a laser range finder and showed that the output of a derivative filter applied to the logarithm of range values is characterized by a pdf similar to that of grayscale images, i.e. with a high kurtosis (high peak, large tail).

The authors of [24] considered the range of a pixel as a RV and modeled the joint pdf of the ranges of two pixels by marginalizing out the RV denoting whether the two pixels belong to the same object or not. When these pixels belong to the same object, the authors model the joint pdf by a large tailed pdf and, otherwise, they assume the statistical independence of the two corresponding range RVs. In our framework, we make the same marginalization, but we derive analytic expressions for the pdfs instead of assuming some parametric large tailed pdfs and learning the parameters of these pdfs on a database of images.

2.1.1 Noise model for range cameras

The image acquisition process usually results in the corruption of the range value of each pixel of an image with noise. A good characterization of the pdf of the noise at a pixel \mathbf{p} is thus helpful to characterize the acquisition process. In range images, the noise can be very different from the noise encountered in luminance images. In images captured by time-of-flight and structured-light cameras, one can observe that the standard deviation of the noise varies across the image. For ToF cameras, experimental and theoretical analyses of the noise have established that it can be modeled by an additive Gaussian noise, whose standard deviation $\sigma_{\text{ToF}}(\mathbf{p})$ at pixel \mathbf{p} is proportional to the inverse of the strength of the signal used to measure the range, commonly called the amplitude, denoted by $A(\mathbf{p})$, and provided by most ToF cameras [19], [27], [30]. In mathematical terms, we have

$$\sigma_{\text{ToF}}(\mathbf{p}) = \sigma_{\text{base}} + \kappa \frac{1}{A(\mathbf{p})}, \quad (1)$$

where σ_{base} is the part of the noise that is independent of the scene, and κ is the factor related to the inverse of the amplitude. The values of σ_{base} and κ may change with the settings of the camera (such as the integration time). We note that this model is valid only for some domain of amplitudes [27]. Outside of this domain, the model in Eq. (1) underestimates the standard deviation of the noise. This occurs when the sensor saturates or when the power level of the return signal is low. The amplitude reflects various noise effects, such as those associated with the reflective properties of the materials and the attenuation of the signal due to the distance between the camera and the scene. We also note that ToF data can produce a bias in the range measurements. However, this bias can be corrected for by the use of appropriate calibration techniques [26].

The Kinect 1 and other SL cameras do not measure the range of the scene directly but, instead, reconstruct range values during a process similar to that of stereoscopic cameras. Khoshelham [20] showed that the noise $\sigma_{\text{SL}}(\mathbf{p})$ at a pixel \mathbf{p} has a Gaussian pdf and that its standard deviation is proportional to the square of the range $z_{\mathbf{p}}$. So, we have

$$\sigma_{\text{SL}}(\mathbf{p}) = \kappa z_{\mathbf{p}}^2, \quad (2)$$

where κ is a constant factor.

By contrast with ToF cameras, the noise for SL cameras is not dependent on scene properties such as the reflectance of materials or the orientation of surfaces. However, the Kinect 1 produces no range data for parts of the scene where the reconstruction method fails.

2.2 Surfaces and edges in range images

By contrast with grayscale and color cameras, range cameras capture geometrical information that is directly linked to the local characteristics of the real world surfaces and their borders. Therefore, it is useful to understand the characteristics of these surfaces and of their edges, and how they are related to the range values in the image. In the literature dedicated to edge detection, one finds models for how real world surfaces and their borders are transformed when imaged by a range camera. For range images, edge detection methods often detect both jump edges and roof

edges. Jump edges correspond to the borders between locally continuous surfaces and occur when one object of the scene occludes another one (occlusion border), while roof edges correspond to the set of real world points where a surface is locally continuous but not differentiable and occur when the orientation of the surface of an object changes abruptly. In the image, the jump and roof edges correspond to discontinuities in the image and its gradient, respectively. In the following, we first deal with how jump and roof edges are perceived by a range camera, and we then focus on methods for edge detection in range images.

SL and ToF cameras do not perceive a jump edge in the same way. First, for a SL camera, the reconstruction performed by the camera guarantees that range values always correspond to real continuous surfaces of the scene. Thus, in the image, a jump edge corresponding to the occlusion border between two real world surfaces can be thought of as being in between the pixels in one or more pairs of neighboring pixels, where one pixel in a pair corresponds to one of the two surfaces and the other pixel in this pair corresponds to the other surface. This jump edge can also be represented by these pairs of neighboring pixels. In other words, the real world border between two surfaces is mapped into the image as a two pixels wide curve.

Second, for a ToF camera, there is no reconstruction step and the range images are measured by the camera. Thus, a jump edge corresponding to an occlusion border between two real world surfaces of the scene can be mapped in the image into one or more transition pixels, called veiled pixels [35], located between the pixels that correspond to each of the two surfaces. The range measured at veiled pixel is a mixture of the ranges corresponding to the two surfaces at their intersection. The ranges of veiled pixels has no real physical meaning. Note that some authors have developed techniques to detect these pixels [2], [35].

A roof edge will always correspond to a point on a continuous surface and its range does not depend on the camera. However, roof edges are more complicated to detect because the standard deviation of the noise of the gradient of the image is higher than the standard deviation of the image itself, and because typical noise removal filters create false roof edges around jump edges.

Table 1 lists and briefly describes state-of-the-art edge detectors specifically designed for range images. They are grouped according to the underlying method used. The table shows that almost all edge detectors for range images can detect jump edges, but that only some can detect roof edges. In this paper, we only consider locally continuous surfaces and jump edges. Furthermore, we see that most edge detectors were designed and evaluated only for a specific kind of range camera. Note that some edge detectors were designed for high precision range scanners, such as LIDARs, that have a very long acquisition time and cannot be used for real time applications.

The next five paragraphs successively describe the five types of edge detectors.

The edge detectors based on derivative operators detect the peaks in the output of the derivative operator applied to the images. Some edge detectors adapt common edge detectors such as the Canny edge detector [7] or the Sobel edge detector to the range image [25], [39]. Some others

detect the edges in a range image by detecting the edges of a corresponding luminance image and then using the range image to categorize the edges into jump, roof, or luminance edges [6], [32]. Indeed, in luminance images, edges usually appear at range or orientation discontinuities. As the sampling grid from typical laser range scanners can deviate from the rectangular grid used in images, Coleman *et al.* [8] designed a method that, based on a finite element framework, adapts differential operators to the irregular sampling of the data.

The edge detectors based on residual analysis estimate the image derivatives by using the difference between the original image and a filtered version thereof [5], [22].

The edge detectors based on model fitting detect edges by estimating how well a given local model, such as a quadratic polynomial [17] or a third-order surface [29], fits the data.

The edge detectors based on machine learning can be categorized according to the machine learning algorithm and the features they use. State of the art detectors have used algorithms such as decision trees [34], support vector machine (SVM) [12], [31], structured forests [9], [13], and convolutional neural networks (deep learning) [38]. Common features are the gradients of the depth computed at different scales and orientations [9], the angle between normals [13], the height of the pixel above the ground, and the angle between the normal at the pixel and the direction of gravity in the image [13], [34], [38].

The edge detectors based on statistical models use statistical tools to determine the edge probability. For example, multiscale Markov random fields have been used to extract jump and roof edges in both range and luminance images [11]. Simple statistics have also been used to estimate the likelihood that a pixel belongs to a jump edge [35].

3 PROBABILISTIC FRAMEWORK

Our framework assumes that the range images are captured by cameras that can be modeled by the projective camera model [14]; this is the case for ToF and SL cameras.

Because surfaces and edges involve, at the very least, two points, the basis of our framework is built around the probabilistic characterization of pairs of pixels in the range image of interest. The pixels in a pair are generically denoted by \mathbf{p} and \mathbf{q} . First, we treat the measured ranges $z_{\mathbf{p}}$ and $z_{\mathbf{q}}$ as the realizations of corresponding RVs, $Z_{\mathbf{p}}$ and $Z_{\mathbf{q}}$. Second, we treat the “belonging or not” of the corresponding real world points to the same surface as the realization of a binary RV $G_{\mathbf{pq}}$. More precisely, the sample space of $G_{\mathbf{pq}}$ consists of the two elementary outcomes $\mathcal{S}_{\mathbf{pq}}$, “the two points belong to the same local and continuous real world surface in the scene”, and $\mathcal{J}_{\mathbf{pq}}$, “the two points straddle a jump edge”. With our definition, we intentionally exclude roof edges since they are ill defined for just two pixels. Note that, by definition, the elementary outcomes $\mathcal{S}_{\mathbf{pq}}$ and $\mathcal{J}_{\mathbf{pq}}$ are complementary, *i.e.* $P(\mathcal{S}_{\mathbf{pq}}) + P(\mathcal{J}_{\mathbf{pq}}) = 1$.

Using this characterization, we decompose the joint pdf $f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}})$ by considering independently the cases where the two corresponding real world points are on the

Table 1

Comparison of a selection of edge detectors for range images. In the sixth column, “ToF” refers to time-of-flight cameras based on frequency modulation, “SL” refers to real time structured light cameras such as the Kinect 1, “range scanners” refers to range cameras with a high precision and a long acquisition time such as LIDARs, and “Synthetic” refers to the process of generating images from a 3D model with a computer.

Edge detectors	Jump edges	Roof edges	Textured required	Noise adaptive	Cameras used in the evaluation	Evaluation	
						Subjective	Objective
Based on derivative filters							
Lejeune <i>et al.</i> [25]	Yes	No	No	Yes	SL, ToF	No	Yes
Schäfer <i>et al.</i> [32]	Yes	If in texture	Yes	No	ToF	No	Yes
Cang Ye <i>et al.</i> [39]	Yes	No	No	No	ToF	No	Yes
Coleman <i>et al.</i> [8]	Yes	Yes	No	No	Synthetic	Yes	Yes
Buch <i>et al.</i> [6]	If in texture	If in texture	Yes	No	SL	No	No
Based on residual analysis							
Boulanger <i>et al.</i> [5]	Yes	Yes	No	No	Range scanners	No	Yes
Krishnapuram <i>et al.</i> [22]	Yes	Yes	No	No	Range scanners	No	Yes
Based on model fitting							
Parvin <i>et al.</i> [29]	Yes	Yes	No	No	Range scanners	No	Yes
Jiang <i>et al.</i> [17]	Yes	Yes	No	No	Range scanners	Yes	Yes
Based on machine learning							
Dollár <i>et al.</i> [9]	Yes	Yes	No	No	Kinect 1	Yes	Yes
Ren <i>et al.</i> [31]	Yes	If in texture	No	No	Kinect 1	Yes	Yes
Gupta <i>et al.</i> [13]	Yes	Yes	No	No	Kinect 1	Yes	Yes
Xie <i>et al.</i> [38]	Yes	Yes	No	No	Kinect 1	Yes	Yes
Silberman <i>et al.</i> [34]	Yes	If in texture	Yes	No	Kinect 1	Yes	Yes
Based on statistical models							
Günsel <i>et al.</i> [11]	Yes	Yes	No	No	Synthetic	No	Yes
Steder <i>et al.</i> [35]	Yes	No	No	No	Range scanners	No	No
Present paper	Yes	Possibly	No	Yes	Kinect 1, ToF	Yes	Yes

same surface ($\mathcal{S}_{\mathbf{p}\mathbf{q}}$) and where they are not ($\mathcal{J}_{\mathbf{p}\mathbf{q}}$) by using the law of total probability:

$$f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}}) = f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}}|\mathcal{S})\pi_{\mathcal{S}} + f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}}|\mathcal{J})\pi_{\mathcal{J}}, \quad (3)$$

where $\pi_{\mathcal{S}} = P(\mathcal{S})$ and $\pi_{\mathcal{J}} = P(\mathcal{J})$ are two probabilities, commonly named priors, and where we dropped the subscript of $\mathcal{S}_{\mathbf{p}\mathbf{q}}$ and $\mathcal{J}_{\mathbf{p}\mathbf{q}}$; see Section 5 for a discussion on the priors.

It is reasonable to assume that (1) the RV $Z_{\mathbf{p}}$ is independent of \mathcal{S} or \mathcal{J} , *i.e.* $f_{Z_{\mathbf{p}}}(z_{\mathbf{p}}) = f_{Z_{\mathbf{p}}}(z_{\mathbf{p}}|\mathcal{S}) = f_{Z_{\mathbf{p}}}(z_{\mathbf{p}}|\mathcal{J})$, and that, (2) the ranges of two pixels straddling an edge are independent, *i.e.* $f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}}|\mathcal{J}) = f_{Z_{\mathbf{p}}}(z_{\mathbf{p}}|\mathcal{J})f_{Z_{\mathbf{q}}}(z_{\mathbf{q}}|\mathcal{J})$. Therefore, we have

$$f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}}|\mathcal{J}) = f_{Z_{\mathbf{p}}}(z_{\mathbf{p}})f_{Z_{\mathbf{q}}}(z_{\mathbf{q}}) \quad (4)$$

$$f_{Z_{\mathbf{p}}, Z_{\mathbf{q}}}(z_{\mathbf{p}}, z_{\mathbf{q}}|\mathcal{S}) = f_{Z_{\mathbf{q}}}(z_{\mathbf{q}}|z_{\mathbf{p}}, \mathcal{S})f_{Z_{\mathbf{p}}}(z_{\mathbf{p}}). \quad (5)$$

In order to derive analytic expressions for these two pdfs, we decompose the measured range RV $Z_{\mathbf{p}}$ into the RV $\bar{Z}_{\mathbf{p}}$ representing the noiseless part of the measured range and the RV $N_{\mathbf{p}}$ representing its additive noise part: $Z_{\mathbf{p}} = \bar{Z}_{\mathbf{p}} + N_{\mathbf{p}}$.

Note that it is possible to learn the pdfs $f_{\bar{Z}_{\mathbf{q}}}(\bar{z}_{\mathbf{q}}|\bar{z}_{\mathbf{p}}, \mathcal{S})$ and $f_{\bar{Z}_{\mathbf{p}}}(\bar{z}_{\mathbf{p}})$, instead of deriving analytic expressions. This permits to tailor the distributions to cameras, scenes, or applications, but requires a representative dataset and a time consuming annotation process. A model based approach, like ours, is more generic as long as the underlying assumptions are verified.

In Section 3.1, we formulate our main hypothesis and challenge it against public databases of range images. In Section 3.2, we derive analytic expressions for $f_{\bar{Z}_{\mathbf{q}}}(\bar{z}_{\mathbf{q}}|\bar{z}_{\mathbf{p}}, \mathcal{S})$ and $f_{\bar{Z}_{\mathbf{p}}}(\bar{z}_{\mathbf{p}})$ for the noiseless part of the measured range. In Section 3.3, we rework these expressions to include the noise and to determine $f_{Z_{\mathbf{q}}}(z_{\mathbf{q}}|z_{\mathbf{p}}, \mathcal{S})$ and $f_{Z_{\mathbf{p}}}(z_{\mathbf{p}})$. In Section

3.4, we derive the probability $P(\mathcal{S}|z_{\mathbf{p}}, z_{\mathbf{q}})$ that, given their ranges, the two real world points corresponding to the pixels \mathbf{p} and \mathbf{q} are on the same real world surface. This is the key for applying the framework to the special case of edge detection.

3.1 Main hypothesis

Our main hypothesis about surfaces in range images is that all surface orientations are equally probable in the physical world. This may not be true for a single scene and camera location and orientation because most surfaces in the real world are aligned with the floor and the walls in a scene. However, if we take into consideration all the possible locations and orientations of the camera for a scene, our hypothesis is acceptable.

To integrate our main hypothesis in the framework, we first formalize the notion of surface orientation. Figure 3 shows the geometrical configuration of interest. The pixels \mathbf{p} and \mathbf{q} in the image plane corresponds to points \vec{P} and \vec{Q} in the real world, respectively. The ranges of \vec{P} and \vec{Q} with respect to the optical center \vec{O} are $\bar{z}_{\mathbf{p}}$ and $\bar{z}_{\mathbf{q}}$, respectively.

Assuming that all surface orientations are equally probable is equivalent to considering that the angle θ between the direction from \vec{Q} to \vec{P} and the reference direction (\mathbf{l}_{om} in Fig. 3) corresponding to the two pixels \mathbf{p} and \mathbf{q} in the image plane is the realization of a RV Θ that is uniformly distributed over $[0, \pi]$, *i.e.* $f_{\Theta}(\theta) = 1/\pi$.

In Fig. 4a, we plotted the histogram of the orientations θ computed for every pair of horizontally and vertically aligned pixels in range images contained in the RGB-D SLAM Dataset [36] (3D Object Reconstruction category, 18913 images captured by a Kinect 1). We choose this particular dataset because images are captured with a moving camera, so that the chance to cover a large spectrum of orientation angles is higher. This histogram has a broad

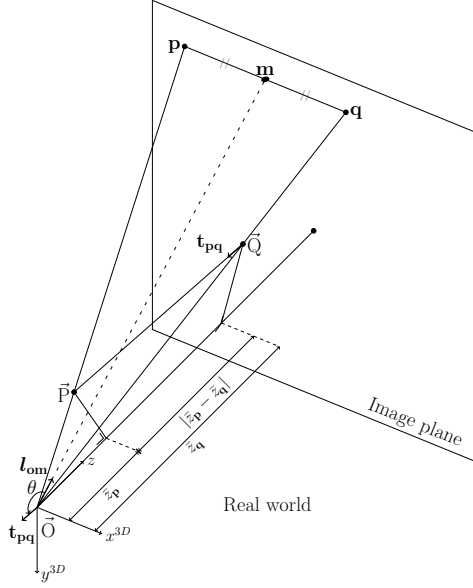


Figure 3. Geometrical configuration showing a pair of pixels p, q in the image plane and their corresponding points P, Q in the real world.

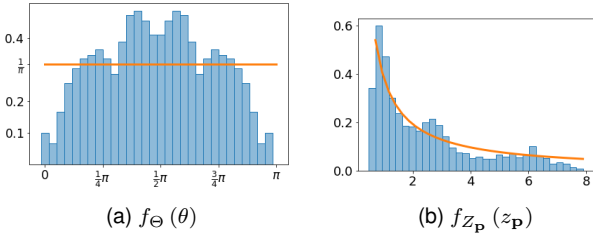


Figure 4. (a) Comparison between the probability density function of the surface orientation $f_{\Theta}(\theta)$ (shown as a continuous line) derived from our assumption that all surface orientations are equally probable in the real world, and the histogram of surface orientations estimated from real world range images from the RGB-D SLAM Dataset [36]. (b) Corresponding comparison for the pdf of range values $f_{Z_p}(z_p)$.

plateau around $\pi/2$ spanning more than half of the possible orientations. Therefore, we believe that our hypothesis of a uniform pdf for the surface orientations is reasonable for arbitrary scenes, and close to the pdfs observed for a benchmarking dataset.

3.2 Determination of $f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})$ and $f_{\bar{z}_p}(\bar{z}_p)$

Appendix A shows that, under the conditions of a projective camera model and equiprobable orientations, the conditional pdf $f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})$ is the Cauchy pdf

$$f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S}) = \frac{1}{\pi \bar{z}_p s_{pq} \left(1 + \left(\frac{\bar{z}_q - \bar{z}_p l_{pq}}{\bar{z}_p s_{pq}} \right)^2 \right)}, \quad (6)$$

where l_{pq} and s_{pq} are respectively the location and scale parameters, as given in Appendix A.4. There, we also make the important observation that l_{pq} and s_{pq} are linearly proportional to z_p , and that they depend on the calibration matrix K of the camera and the coordinates of the pixels p and q .

The Cauchy pdf being fat-tailed, Eq. (6) shows that large range differences on a surface are possible and that they do

not necessarily imply a jump edge. Moreover, because the mean and variance of a Cauchy RV are undefined [28], any surface model based on them is flawed. In particular, trying to estimate them from a database of images is questionable, to say the least. This emphasizes the importance of having an analytic expression for the pdf of range values on surfaces. To the best of our knowledge, this is the first time that an analytic expression is given for $f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})$. In [15], $f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})$ is estimated experimentally by fitting range values to a pdf that is “highly concentrated around 0”.

Let us now discuss $f_{\bar{z}_p}(\bar{z}_p)$. Without any knowledge about the content of the scene, a reasonable form for $f_{\bar{z}_p}(\bar{z}_p)$ is that of a uniform pdf. In our case, because of the assumptions made to derive $f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})$, we have a different pdf. Indeed, Bayes’ theorem leads to

$$\frac{f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})}{f_{\bar{z}_p}(\bar{z}_p|\bar{z}_q, \mathcal{S})} = \frac{f_{\bar{z}_q}(\bar{z}_q)}{f_{\bar{z}_p}(\bar{z}_p)}. \quad (7)$$

From Eq. (6), we can see that $f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})$ differs from $f_{\bar{z}_p}(\bar{z}_p|\bar{z}_q, \mathcal{S})$. Algebraic developments, confirmed with the help of the Mathematica software package, lead to

$$\frac{f_{\bar{z}_q}(\bar{z}_q|\bar{z}_p, \mathcal{S})}{f_{\bar{z}_p}(\bar{z}_p|\bar{z}_q, \mathcal{S})} = \frac{\bar{z}_p}{\bar{z}_q}. \quad (8)$$

Thus, $f_{\bar{z}_p}(\bar{z}_p)$ must take the form of a Reciprocal pdf, *i.e.* $f_{\bar{z}_p}(\bar{z}_p) = K \bar{z}_p^{-1}$, where K is a normalizing constant to be determined. For positive ranges comprised between z_{min} and z_{max} , we have

$$f_{\bar{z}_p}(\bar{z}_p) = \begin{cases} \frac{1}{\ln z_{max} - \ln z_{min}} \frac{1}{\bar{z}_p} & \text{if } \bar{z}_p \in [z_{min}, z_{max}], \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where $\ln x$ denotes the natural logarithm of x . The ranges z_{min} and z_{max} can be set according to the camera specifications or computed for each range image.

While the pdf for $f_{\bar{z}_p}(\bar{z}_p)$ is identical for every pixel in the image, the shape of the Reciprocal pdf means that smaller ranges are more likely than large range values. This is a consequence of the assumption that the orientations of real world surfaces in the scene with respect to the line of sight are equally probable. We show in Fig. 4b that this pdf corresponds reasonably well to real world data.

Other authors have provided other expressions for $f_{\bar{z}_p}(\bar{z}_p)$ but under different assumptions. In [21], Konishi *et al.* learn the pdf over a database by computing a histogram of the ranges from the neighborhood around the considered pixel. In [15], Huang *et al.* use knowledge on the content of the scene (forest scenes) to determine an analytic expression for the pdf.

Interestingly, the inverse of a RV with a Cauchy pdf also has a Cauchy pdf, and similarly for a RV with a Reciprocal pdf. We explain in Section 4.1 that this is particularly useful when dealing with planar surfaces, where using the inverse range is more appropriate, as shown in [37].

3.3 Determination of $f_{z_q}(z_q|z_p, \mathcal{S})$ and $f_{z_p}(z_p)$

Until now, we have ignored the noise affecting range values. As noted in Section 2.1.1, this noise can be modeled as an additive Gaussian noise, expressed as $N_p \sim \mathcal{N}(0, \sigma_p^2)$ for a pixel p . Appendix A.5 shows that, if we combine the noise

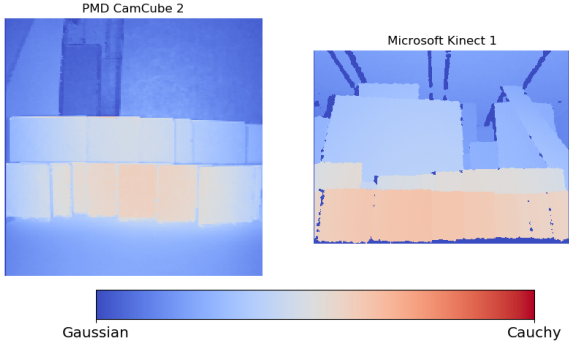


Figure 5. Each image shows, in false colors, and for the indicated range camera, the value, at each pixel, of the parameter η that appears in Eq. (47) of Appendix A.5 and indicates the weight of the Cauchy pdf when one represents the Voigt pdf of $f_{z_q}(z_q|z_p, \mathcal{S})$ as a linear combination of a Cauchy pdf and a Gaussian pdf. The value shown at a pixel \mathbf{q} shows the value of η with \mathbf{p} being immediately to the right of \mathbf{q} . The value of η varies across each image according to the local noise level and the range at each pixel.

affecting \mathbf{p} and \mathbf{q} by taking $\sigma^2 = \sigma_p^2 + \sigma_q^2$, the conditional pdf $f_{z_q}(z_q|z_p, \mathcal{S})$ for noisy range values is the convolution of a Cauchy pdf and a Gaussian pdf, which results in a Voigt pdf. If the standard deviation of the noise is much larger than the scale parameter of the Cauchy pdf, we can approximate $f_{z_q}(z_q|z_p, \mathcal{S})$ by a Gaussian pdf. Conversely, if it is much smaller, we can approximate $f_{z_q}(z_q|z_p, \mathcal{S})$ by a Cauchy pdf. It is important to understand that the shape of $f_{z_q}(z_q|z_p, \mathcal{S})$ is non-uniform over a range image, as illustrated in Fig. 5 for two range images.

Likewise, $f_{z_p}(z_p)$ is the convolution of $f_{\bar{z}_p}(\bar{z}_p)$ and a Gaussian pdf. It is approximately equal to $f_{\bar{z}_p}(\bar{z}_p)$ since the domain of possible range values is usually several orders of magnitude larger than the standard deviation of the noise.

3.4 Determination of $P(\mathcal{S}|z_p, z_q)$

We now determine the conditional probability $P(\mathcal{S}|z_p, z_q)$ that, given the ranges of the two pixels \mathbf{p} and \mathbf{q} , their corresponding real world points are on the same real world surface. A variant of Bayes theorem [28] gives

$$P(\mathcal{S}|z_p, z_q) = \frac{f_{z_q}(z_q|z_p, \mathcal{S}) \pi_{\mathcal{S}}}{f_{z_q}(z_q|z_p, \mathcal{S}) \pi_{\mathcal{S}} + f_{z_q}(z_q) \pi_{\mathcal{J}}}. \quad (10)$$

One sees that $f_{z_q}(z_q|z_p, \mathcal{S})$ and $f_{z_q}(z_q)$ both contribute to $P(\mathcal{S}|z_p, z_q)$. The probability of the outcome \mathcal{J} is $P(\mathcal{J}|z_p, z_q) = 1 - P(\mathcal{S}|z_p, z_q)$, since the two events are complementary.

4 PROBABILISTIC MODEL FOR SURFACES IN THE LOCAL NEIGHBORHOOD OF A PIXEL

So far, we have considered the geometrical relationship between two real world points to determine an expression for the joint pdf $f_{z_p, z_q}(z_p, z_q)$ and to derive the probability that the real world points corresponding to the two pixels \mathbf{p} and \mathbf{q} are on the same surface ($\mathcal{S}_{\mathbf{p}\mathbf{q}}$) or straddle an edge ($\mathcal{J}_{\mathbf{p}\mathbf{q}}$). However, we can enhance our probabilistic model by using more than two pixels. Because we intend to use the framework for edge detection, we consider pixels that are

collinear in the range image. The purpose is to better identify the geometrical configuration (surface or edge) of the real world points corresponding to \mathbf{p} and \mathbf{q} by considering additional pixels along the line going through them. Below, we simplify the notation by dropping the pdf subscripts, so that we write, e.g., $f(z_p, z_q)$.

Let us consider the N aligned and ordered pixels $\mathbf{p}_1, \dots, \mathbf{p}_N$ defined as $\mathbf{p}_i = \mathbf{p}_1 + a_i(\mathbf{p}_N - \mathbf{p}_1)$ with $0 < a_i < a_{i+1} < 1$. The pdf $f(z_1, \dots, z_N)$ of the range value of these N pixels can be expressed using the law of total probability over the set of disjoint events representing the geometrical configurations $c_i = \{\mathcal{S}_{\mathbf{p}_i \mathbf{p}_{i+1}}, \mathcal{J}_{\mathbf{p}_i \mathbf{p}_{i+1}}\}$, $i = 1, \dots, N-1$, between each pair of real world points corresponding to successive pixels,

$$f(z_1, \dots, z_N) = \sum_{c_1} P(c_1) \left(\sum_{c_2} P(c_2) \left[\dots \sum_{c_{N-1}} P(c_{N-1}) f(z_1, \dots, z_N | c_1, \dots, c_{N-1}) \right] \right). \quad (11)$$

There are in fact 2^{N-1} possible geometrical configurations for all the real world points corresponding to the N pixels. Consequently, there also are 2^{N-1} terms in the expression for $f(z_1, \dots, z_N)$.

We can use the independence of the ranges corresponding to different real world surfaces to decompose the joint pdf $f(z_1, \dots, z_N | c_1, \dots, c_{N-1})$. Specifically, when the geometrical configuration c_i is $\mathcal{J}_{\mathbf{p}_i \mathbf{p}_{i+1}}$, we have $f(z_1, \dots, z_N | c_1, \dots, c_{N-1}) = f(z_1, \dots, z_i | c_1, \dots, c_i) f(z_{i+1}, \dots, z_N | c_{i+1}, \dots, c_{N-1})$. By applying this rule to all jump edge configurations, the pdf $f(z_1, \dots, z_N | c_1, \dots, c_{N-1})$ is split into pdfs involving ranges corresponding to the same real world surface. For example, for $N = 3$, the marginalization of $f(z_1, z_2, z_3)$ has four terms,

$$f(z_1, z_2, z_3) = f(z_1, z_2, z_3 | \mathcal{S}_{\mathbf{p}_1 \mathbf{p}_3}) P(\mathcal{S}_{\mathbf{p}_1 \mathbf{p}_2}) P(\mathcal{S}_{\mathbf{p}_2 \mathbf{p}_3}) + f(z_1, z_2 | \mathcal{S}_{\mathbf{p}_1 \mathbf{p}_2}) f(z_3) P(\mathcal{S}_{\mathbf{p}_1 \mathbf{p}_2}) P(\mathcal{J}_{\mathbf{p}_2 \mathbf{p}_3}) + f(z_1) f(z_2, z_3 | \mathcal{S}_{\mathbf{p}_2 \mathbf{p}_3}) P(\mathcal{J}_{\mathbf{p}_1 \mathbf{p}_2}) P(\mathcal{S}_{\mathbf{p}_2 \mathbf{p}_3}) + f(z_1) f(z_2) f(z_3) P(\mathcal{J}_{\mathbf{p}_1 \mathbf{p}_2}) P(\mathcal{J}_{\mathbf{p}_2 \mathbf{p}_3}). \quad (12)$$

In Section 3.1, we obtained an expression for $f(z_i, z_{i+1} | \mathcal{S}_{\mathbf{p}_i \mathbf{p}_{i+1}})$ for subsets of two pixels. When the subset contains more than two pixels, we have to make one additional assumption about the underlying surface. In Section 4.1, we provide an analytic expression for $f(z_1, z_2, z_3 | \mathcal{S}_{\mathbf{p}_1 \mathbf{p}_3})$ for the case where all surfaces in the scene are planar surfaces, which is a reasonable assumption locally. In Section 4.2, we derive the probability $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_p, z_q, z_r)$ that the real world points corresponding to the pixels \mathbf{p} and \mathbf{q} are on the same surface given their range and the range of another pixel \mathbf{r} located on the line joining \mathbf{p} and \mathbf{q} .

4.1 Determination of $f(z_1, \dots, z_N | \mathcal{S}_{\mathbf{p}_1 \mathbf{p}_N})$

While the real world coordinates of the points in a plane are related by a linear equation, there is no linear relationship between the coordinates of the corresponding pixels and

their ranges. However, the inverse of the ranges are linearly dependent on the pixel coordinates, as shown in [37]. Thus, given two pixels \mathbf{p}_i and \mathbf{p}_j and their noiseless ranges \bar{z}_i and \bar{z}_j , the range \bar{z}_k of any pixel \mathbf{p}_k (including \mathbf{p}_i and \mathbf{p}_j), such that $\mathbf{p}_k = \mathbf{p}_i + a_{ijk}(\mathbf{p}_j - \mathbf{p}_i)$, where a_{ijk} defines the position of the pixel \mathbf{p}_k with respect to \mathbf{p}_i and \mathbf{p}_j , is related to the other two ranges by $\bar{z}_k^{-1} = \bar{z}_i^{-1} + a_{ijk}(\bar{z}_j^{-1} - \bar{z}_i^{-1})$.

For the set of N pixels \mathbf{p}_k , $k = 1, \dots, N$, we can write the N such equations in the following matrix form

$$\bar{\mathbf{z}}^{-1} = \mathbf{A}\bar{\boldsymbol{\beta}}, \quad (13)$$

where $\bar{\mathbf{z}}^{-1} = (\bar{z}_1^{-1} \dots \bar{z}_i^{-1} \dots \bar{z}_N^{-1})^T$, $\bar{\boldsymbol{\beta}} = (\bar{z}_1^{-1} \bar{z}_N^{-1})^T$, and \mathbf{A} is the $N \times 2$ matrix

$$\mathbf{A} = \begin{pmatrix} \vdots & \vdots \\ 1 - a_{1Ni} & a_{1Ni} \\ \vdots & \vdots \end{pmatrix}. \quad (14)$$

It can be shown that the joint pdf of the noise on the N inverse ranges \mathbf{z}^{-1} is a multivariate Gaussian pdf with zero mean vector and an $N \times N$ diagonal variance matrix $\boldsymbol{\Sigma} = \text{diag}([\sigma_1/z_1^2]^2 \dots [\sigma_N/z_N^2]^2)$. By combining the pdf of the noise with Eq. (13), it can be shown that the pdf of the inverse noisy ranges of N aligned and ordered pixels corresponding to the same planar surface can be approximated by

$$f(\mathbf{z}^{-1} | \mathcal{S}_{\mathbf{p}_1 \mathbf{p}_N}) = \frac{(2\pi)^{-N/2} \sqrt{|\boldsymbol{\Sigma}_\xi^{-1}|_{11}^{-1} [\boldsymbol{\Sigma}_\xi]_{22}}}{[\ln z_{\max} - \ln z_{\min}] \widehat{z}_N^{-1}} G(\mathbf{z}^{-1}; \mathbf{A}\xi, \boldsymbol{\Sigma}) V\left(\widehat{z}_1^{-1}; \widehat{z}_N^{-1} l_{1N}, \sqrt{|\boldsymbol{\Sigma}_\xi^{-1}|_{11}^{-1} + [\boldsymbol{\Sigma}_\xi]_{22}}, \widehat{z}_N^{-1} s_{1N}\right), \quad (15)$$

where $\boldsymbol{\Sigma}_\xi = (\mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{A})^{-1}$ is the covariance matrix of the estimator, $[\boldsymbol{\Sigma}_\xi]_{ij}$ is the element at row i and column j of $\boldsymbol{\Sigma}_\xi$, $G(x; \mu, \sigma)$ is the Gaussian pdf at x with mean μ and standard deviation σ , $\xi = (\widehat{z}_1^{-1} \widehat{z}_N^{-1})^T = (\mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{z}^{-1}$ is the maximum likelihood estimator of the ranges at pixel \mathbf{p}_1 and \mathbf{p}_N given all N ranges, and $V(\widehat{z}_1^{-1}; \dots)$ is the Voigt pdf as defined in Appendix A.5. We note that $G(\mathbf{z}^{-1}; \mathbf{A}\xi, \boldsymbol{\Sigma})$ is the maximum likelihood of the planar model given the N ranges. Eq. (15) is an important result because using more than two points allows one to increase the precision and robustness.

By performing a standard change of RV, we obtain the pdf for $\mathbf{z} = (z_1 \dots z_i \dots z_N)^T$ (instead of for \mathbf{z}^{-1}),

$$f(\mathbf{z} | \mathcal{S}) = \left(\prod_{i=1}^N \frac{1}{z_i^2} \right) f(\mathbf{z}^{-1} | \mathcal{S}). \quad (16)$$

4.2 Determination of $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})$

In the previous section, we have obtained an expression for the joint pdf for the ranges of more than two collinear pixels. In the case of a planar surface model, this allows one to calculate the probability of $\mathcal{S}_{\mathbf{p}\mathbf{q}}$ by considering more than two ranges, which should prove more robust. For example, if we consider a third pixel \mathbf{r} , we have

$$P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}}) = \frac{f(z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}} | \mathcal{S}_{\mathbf{p}\mathbf{q}}) P(\mathcal{S}_{\mathbf{p}\mathbf{q}})}{f(z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})}, \quad (17)$$

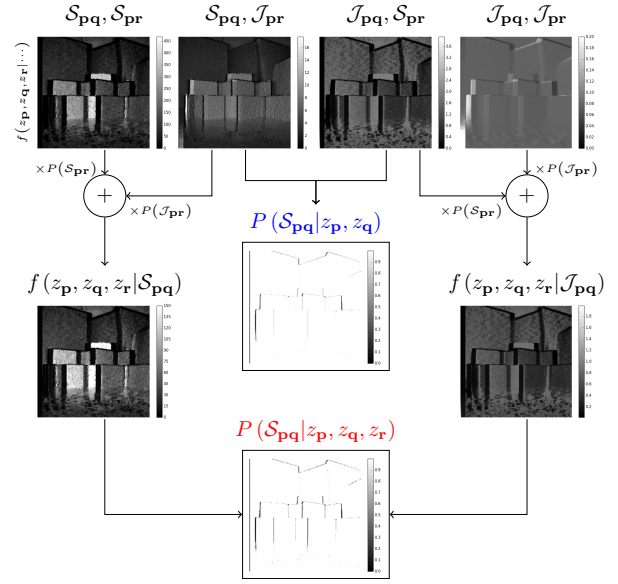


Figure 6. Illustration of the procedure for combining the pdfs $f(z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}} | \dots)$ of horizontally aligned pixels to produce the probability map $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_{\mathbf{p}}, z_{\mathbf{q}})$ based on two ranges $z_{\mathbf{p}}$ and $z_{\mathbf{q}}$, and the probability map $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})$ based on three ranges $z_{\mathbf{p}}$, $z_{\mathbf{q}}$, and $z_{\mathbf{r}}$. The figure shows that adding a third range leads to more edges being detected. This is the basis for our edge detector.

where $f(z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}} | \mathcal{S}_{\mathbf{p}\mathbf{q}})$ is obtained using the law of total probability over the outcomes $\mathcal{S}_{\mathbf{q}\mathbf{r}}$ and $\mathcal{J}_{\mathbf{q}\mathbf{r}}$, and $f(z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})$ over the four configurations $(\mathcal{S}_{\mathbf{p}\mathbf{q}}, \mathcal{S}_{\mathbf{q}\mathbf{r}})$, $(\mathcal{J}_{\mathbf{p}\mathbf{q}}, \mathcal{S}_{\mathbf{q}\mathbf{r}})$, $(\mathcal{S}_{\mathbf{p}\mathbf{q}}, \mathcal{J}_{\mathbf{q}\mathbf{r}})$, and $(\mathcal{J}_{\mathbf{p}\mathbf{q}}, \mathcal{J}_{\mathbf{q}\mathbf{r}})$, as explained in Section 4. In other words, we utilize the information provided by $z_{\mathbf{r}}$ for some pixel \mathbf{r} in addition to that provided by \mathbf{p} and \mathbf{q} to improve the $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | \dots)$. Figure 6 illustrates the decomposition of $f(z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})$ using Eq. (12) and how its terms can be combined to produce the probability maps $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})$ and $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | z_{\mathbf{p}}, z_{\mathbf{q}})$. Observe that the addition of a third range provides additional edges.

In the following section, we present an algorithm that is based on our probabilistic framework and that detects edges in range images. This algorithm illustrates the underlying mechanisms and proves the usefulness of our model for a practical task.

5 APPLICATION TO EDGE DETECTION

We can perform edge detection by thresholding the surface probability $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | \dots)$ for each pair of neighboring pixels \mathbf{p} and \mathbf{q} using

$$\mathcal{S}_{\mathbf{p}\mathbf{q}} \quad \text{if } P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | \dots) > \tau, \quad (18)$$

$$\mathcal{J}_{\mathbf{p}\mathbf{q}} \quad \text{otherwise}, \quad (19)$$

where τ is a detection threshold. In practice, at each pixel \mathbf{p} in the image, we say that there is an edge at \mathbf{p} if $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | \dots) \leq \tau$ for its horizontal neighbor $\mathbf{q} = \mathbf{p} + (1, 0)$ or its vertical neighbor $\mathbf{q} = \mathbf{p} + (0, 1)$. The resulting edge detector is referred to, below, as the Probabilistic Edge Detector (PED).

We define three variants of the PED depending on the number of pixels considered for estimating the surface probability $P(\mathcal{S}_{\mathbf{p}\mathbf{q}} | \dots)$:

- PED 0, which uses the two pixels \mathbf{p} and \mathbf{q} , and their ranges $z_{\mathbf{p}}$ and $z_{\mathbf{q}}$ to determine $P(\mathcal{S}_{\mathbf{pq}}|z_{\mathbf{p}}, z_{\mathbf{q}})$ defined in Eq. (10),
- PED 1, which uses one additional pixel and its range, and
- PED 2, which uses two additional pixels and their ranges.

For PED 2, the two additional pixels are symmetrically taken at the same distance from the center of mass of the two pixels \mathbf{p} and \mathbf{q} , on each side of the line joining \mathbf{p} and \mathbf{q} . We denote by \mathbf{o} the pixel that is on the side of \mathbf{p} , and by \mathbf{r} the pixel that is on the side of \mathbf{q} . In mathematical terms, they are defined by $\mathbf{o} = \mathbf{p} + k(\mathbf{p} - \mathbf{q})$ and $\mathbf{r} = \mathbf{q} + k(\mathbf{q} - \mathbf{p})$, where k is a distance parameter. This allows the edge detector to leverage the information provided by the planar surface model even when there is an edge on only one side of the line passing through \mathbf{p} and \mathbf{q} . For example, if there is a *jump edge* between \mathbf{o} and \mathbf{p} , the range $z_{\mathbf{r}}$ still contributes to represent the local surface containing the real world points corresponding to the pixels \mathbf{p} , \mathbf{q} , and \mathbf{r} . The distance k defines the size of the spatial support for the surface model. This size is critical because, on the one hand, it has to be small to accurately represent the local surface while, on the other hand, it must be large enough to reduce the effect of the noise and to increase the confidence in the planar surface model.

For PED 1, we choose as the additional pixel either \mathbf{o} or \mathbf{r} , favoring the point that is most likely on the same surface as \mathbf{p} and \mathbf{q} . In practice, in order to avoid computing both the probabilities $P(\mathcal{S}_{\mathbf{pr}}|z_{\mathbf{p}}, z_{\mathbf{q}}, z_{\mathbf{r}})$ and $P(\mathcal{S}_{\mathbf{oq}}|z_{\mathbf{o}}, z_{\mathbf{p}}, z_{\mathbf{q}})$, a good rule of thumb is to choose the pixel the range of which is the closest to $z_{\mathbf{p}}$ and $z_{\mathbf{q}}$. Similarly to PED 2, this choice allows the detector to leverage the information provided by the planar surface model even when there is an edge on only one side of \mathbf{p} and \mathbf{q} .

By construction, our technique detects edges that are at least two pixels thick, with one pixel on each side of the discontinuity. The detected edges are even thicker if there is a veiled pixel at the discontinuity. However, in some applications and when needed, it is possible to interpret the role of a pair of edge pixels in different terms. Indeed, in a pair of edge pixels, the pixel with the smallest range distance most likely belongs to the inner border of a foreground object, and the other to the background of the scene. Since such interpretations are scene dependent, we ignore them here.

PED 0 is directly based on two parameters, *i.e.* the prior $\pi_{\mathcal{J}} = P(\mathcal{J}_{\mathbf{pq}})$ and the detection threshold τ . In addition, PED 1 and PED 2 also involve the priors $P(\mathcal{J}_{\mathbf{op}})$ and $P(\mathcal{J}_{\mathbf{qr}})$, which can be estimated as follows. By definition, $P(\mathcal{J}_{\mathbf{op}}) = 1 - P(\mathcal{S}_{\mathbf{op}})$. To estimate $P(\mathcal{S}_{\mathbf{op}})$, we should remember that the pixels \mathbf{o} and \mathbf{p} are separated by a distance k , so that there should be no edge between the $k - 1$ pairs of adjacent pixels in between \mathbf{o} and \mathbf{p} . Therefore, $P(\mathcal{S}_{\mathbf{op}}) = (1 - \pi_{\mathcal{J}})^{k-1}$, and

$$P(\mathcal{J}_{\mathbf{op}}) = P(\mathcal{J}_{\mathbf{qr}}) = 1 - (1 - \pi_{\mathcal{J}})^{k-1}. \quad (20)$$

The value of the edge prior $\pi_{\mathcal{J}}$ depends on the camera type, the resolution of the range image, and the kind of scene captured by the camera. It should thus be chosen or estimated accordingly.

The threshold τ indicates how well the range values must fit the surface model in order to tell whether the real world points corresponding to the pixels \mathbf{p} and \mathbf{q} are on the same surface. Using a larger τ value increases the number of true edges, but also the number of edges falsely declared to be true edges (false positives). A proper strategy to determine τ consists in maximizing the accuracy while keeping the number of false positives small.

6 EXPERIMENTAL RESULTS AND COMPARISON

In this section, we present two series of experiments. First, we analyze the performance of our three PED edge detectors defined in Section 5. Second, we compare them to a selected set of state of the art edge detectors.

To evaluate the performance of all the edge detectors considered, we follow the methodology of Arbelaez *et al.* [3], which uses the precision-recall framework and defines three indicators: (1) the ODS, which is the best F score (harmonic mean of the precision and the recall) on the dataset using the same set of parameters, (2) the OIS, which is the aggregate F score obtained for the best set of parameters in each image, and (3) the average precision AP, also called the area under the precision-recall curve. The evaluation indicators are made invariant to the width of the detected edges by applying a standard edge thinning algorithm. When computing the performance values, a small localization error of the detected edge with respect to the ground truth is tolerated to make the results more robust to an inexact localization in the ground truth.

The series of experiments related to our PEDs use a custom made dataset, hereafter referred to as the JUMP dataset, containing manually annotated ground truth jump edges for range images captured by three different range cameras having different properties:

- Kinect 1 (structured light, resolution of 640×480 pixels),
- PMD CamCube 2.0 (single modulation frequency time-of-flight, 204×204 pixels),
- Microsoft Kinect 2 (multiple modulation frequency time-of-flight, 512×424 pixels).

We captured six images for each of these cameras. The scenes captured in this dataset were designed to have a large variety of jump edges. We preprocessed the images from ToF cameras using a 3×3 median filter to remove a significant “salt and pepper” noise and setting to zero the pixels with amplitudes outside the validity zone (see Section 2.1.1), creating holes similar to those of the Kinect 1.

For the comparative study of our PEDs with the set of state of the art detectors, we use the well established NYU Depth Dataset (NYUD) [34]. We note that, while it is often used, this dataset is not ideal for the evaluation of jump edge detectors on range images. First, the edges delineate semantic objects in the image and they may correspond to a jump edge, a roof edge, or an edge in the RGB image. Second, the dataset was captured using a single type of range camera.

6.1 Analysis of the parameters of PEDs

The PEDs have five parameters: PED# (PED 1, 2 or 3), noise factor κ , distance k , edge prior $\pi_{\mathcal{J}}$, and threshold τ . The aim

Table 2
Results of our edge detector for each camera used to produce our JUMP dataset.

		ODS	OIS	AP
Kinect 1	PED 0	.955	.964	.983
	PED 1	.959	.970	.983
	PED 2	.956	.968	.982
Kinect 2	PED 0	.946	.958	.985
	PED 1	.959	.967	.985
	PED 2	.966	.971	.994
CamCube	PED 0	.862	.887	.918
	PED 1	.876	.895	.932
	PED 2	.880	.902	.937

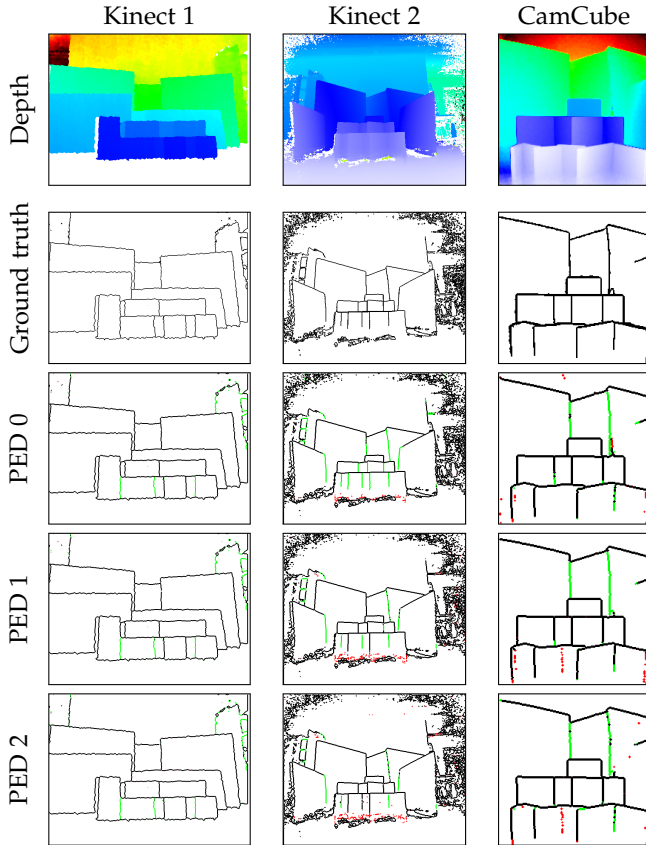


Figure 7. Edge detection results on our JUMP dataset. Black, red, and green pixels represents the true positives, false positives, and false negatives, respectively.

of this section is to analyze how these parameters influence the performance of the detector and to provide a good set of default parameters.

Table 2 reports the best performances obtained by extensively searching the parameter space with the three versions of the PED on our JUMP dataset. Figure 7 shows a few edge detection results on this dataset.

Figure 8 shows a sensitivity analysis of each parameter of the PED on the images of the Kinect 1 and Kinect 2. When one parameter is tested, the value of each of the other parameters is set to the value that gives the best ODS score as shown in Table 2. We now analyze the results of Fig. 8 for each parameter.

PED 1 and PED 2 (*i.e.* our edge detectors that considers a planar surface model for a set of respectively three and four pixels) outperforms PED 0 (*i.e.* our edge detector that only

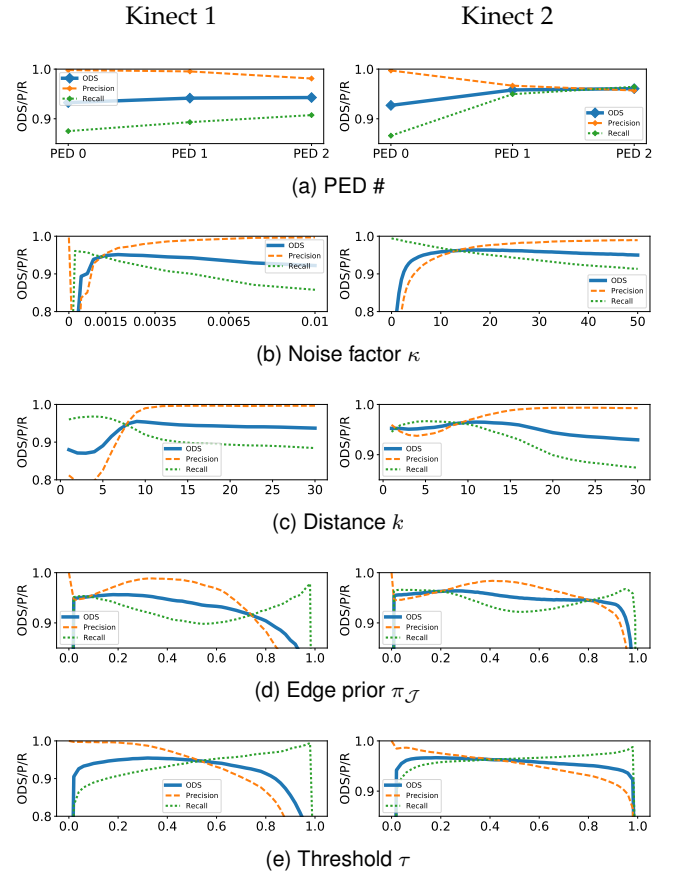


Figure 8. Influence of the parameters of the Probabilistic Edge Detector (PED) on the performance (ODS, recall, and precision) for the Kinect 1 and Kinect 2 images from our custom JUMP dataset.

considers two pixels) for the Kinect 2 and PMD CamCube 2.0. This shows the importance of considering the ranges of enough pixels to build a more precise surface model. For the Kinect 1, the use of a surface model has a smaller impact on the performance on our custom dataset but gives a noticeable boost on the NYUD (see Table 3). We observe that the difference in performance between PED 1 and PED 2 is small on our JUMP dataset. This may be because of how we choose the third pixel in PED 1. However, we still expect PED 2 to be more efficient on scenes where there are surfaces that are very steep with respect to the camera.

For the noise factor κ , the best performances are reached for values that are very close to the values we estimated, experimentally, with the cameras ($\kappa_{\text{Kinect 1}} \approx 0.0015$ and $\kappa_{\text{Kinect 2}} \approx 12$).

The best distance k at which to retrieve the additional pixels is around 8 for the two cameras. A smaller value increases the influence of the noise in Eq. (15) and a larger value will increase the probability that the additional pixels are not on the same surface.

The edge prior $\pi_{\mathcal{J}}$ has little influence on the performance as long it is between 0.01 and 0.3. Its value on our JUMP dataset is 0.028 for the Kinect 1 and 0.037 for the Kinect 2.

The threshold τ gives the best performance for values around 0.5. It allows one to specify the trade off between the precision and the recall: small values of τ give a high precision and a low recall, while, for large values of τ , the effect is opposite.

Table 3

Best performance indicators obtained for each of the 12 edge detectors on the NYU Depth Dataset. For the detectors based on machine learning (Dollar, Gupta, and Xie), we duplicated the analysis by considering a complete model learned on the NYU Depth Dataset (“NYUD model”) or a jump edge model learned on the Kinect 1 images of our custom JUMP dataset (“JUMP model”).

	ODS	OIS	AP
Jiang [17]	.447	.487	.574
Lejeune [25]	.537	.568	.562
Sobel	.490	.519	.560
Dollar [9] (NYUD model)	.642	.655	.660
Gupta [13] (NYUD model)	.640	.650	.660
Xie [38] (NYUD model)	.682	.695	.702
Dollar [9] (JUMP model)	.519	.529	.389
Gupta [13] (JUMP model)	.541	.561	.507
Xie [38] (JUMP model)	.341	.349	.123
PED 0	.502	.504	.705
PED 1	.554	.577	.614
PED 2	.541	.569	.595

Our conclusion is that a good default choice would be to take the PED 1 with a noise factor measured experimentally for the camera, and the following parameter set: $k = 8$, $\pi_{\mathcal{G}} = 0.1$, and $\tau = 0.5$.

The processing times of the PED 0, 1, and 2 on a Microsoft Kinect 1 image (VGA resolution) are respectively 30 ms, 115 ms, and 230 ms on a Intel(R) Core(TM) i7-4500U for the C++ implementation provided at www.telecom.ulg.ac.be/range-model.

6.2 Comparative study

Table 3 gives the performance of the three versions of PED and the edge detectors of Jiang [17] (jump edge only), Lejeune [25], Sobel, Dollar [9], Gupta [13], and Xie [38] on NYUD. We use the same matching tolerance of 1.1% of the size of the image diagonal as done in [9], [13], [38]. Figure 9 shows edge detection results for a representative selection of images and detectors.

The results of Table 3 were obtained for the three PEDs, Jiang, Lejeune, and Sobel by extensively searching each of their parameter space to reach the maximum scores. The Jiang and Sobel edge detector were modified to take into account the per-pixel noise as discussed in Section 2.1.1. The results for the machine learning based edge detectors (Dollar, Gupta, and Xie) are given for a full model (jump/roof edges) learned on NYUD (depth only, train/test separation identical to that of [13]) and for a jump edge model learned on the Kinect 1 images of our JUMP dataset.

We can see that the machine learning based edge detectors with a full model learned on NYUD outperform all other detectors. Indeed, they detect a large number of roof edges present in the ground truth that the other detectors do not detect. However, when machine learning based detectors are trained on a jump edge model only, their performance is below that of PED 1. It also appears that the Xie edge detector is subject to underfitting or overfitting since it achieves poor results with the jump edge model. The Jiang and Sobel edge detectors are the worst performing detectors in this evaluation.

With respect to the PEDs, we see that PED 1 and PED 2 always perform better than PED 0. This shows the importance of considering some surface model on a neighborhood to achieve good detection. We think that additional improvements could be made by considering more pixels and/or using a double thresholding scheme such as in [7], [25].

In summary, we see that jump edge detectors using a model based approach built on our framework are efficient, especially when we consider more than two pixels and a surface model. Moreover, our framework can be used with other range cameras with little effort, whereas machine learning based approaches require the time consuming task of building a representative dataset for each type of camera.

7 CONCLUSION

Range images directly reflects the geometrical structure of the corresponding real world scene. This makes the interpretation and the analysis of this scene from range image more straightforward than from classic intensity images. In this work, we studied the link between the range data and the image acquisition process to obtain a powerful probabilistic framework for processing range images.

First, we derived an analytic expression for the pdf of range values when the corresponding points of the scene belong to the same surface. Using the assumption that all surface orientations are equally probable, we established that range values follow a Voigt pdf. This conclusion is consistent with previous empirical observations for the pdf of range values, *i.e.* a high peak and a large tail. We extended our probabilistic model to characterize the probability that sets of collinear pixels are sampled from the same planar surface in the scene. We proved that the expression for the pdf of the ranges of a set of collinear pixels is the maximum likelihood of the planar model weighted by a Voigt pdf.

Second, we designed a jump edge detector to illustrate an application of our framework and to analyze the influence of its parameters in a real use case. Our Probabilistic Edge Detector (PED) has good performances compared to state of the art detectors, including machine learning based detectors. These results are a consequence of the intrinsic capability of our probabilistic framework to adapt the detector to the local noise and to consider a more complex surface model while rejecting regions that provide a poor description of the local surface.

More generally, we developed our framework for range images to express the notion of edges and surfaces in a probabilistic manner. We think that our model based framework could be used for various image processing tasks on range images such as noise filtering and segmentation, and in association with machine learning based methods by building probabilistic features.

Our custom, annotated JUMP dataset, source code, and illustrative results are publicly available at www.telecom.ulg.ac.be/range-model.

ACKNOWLEDGMENTS

This work was funded by the European Regional Development Fund program of the Walloon Region of Belgium.

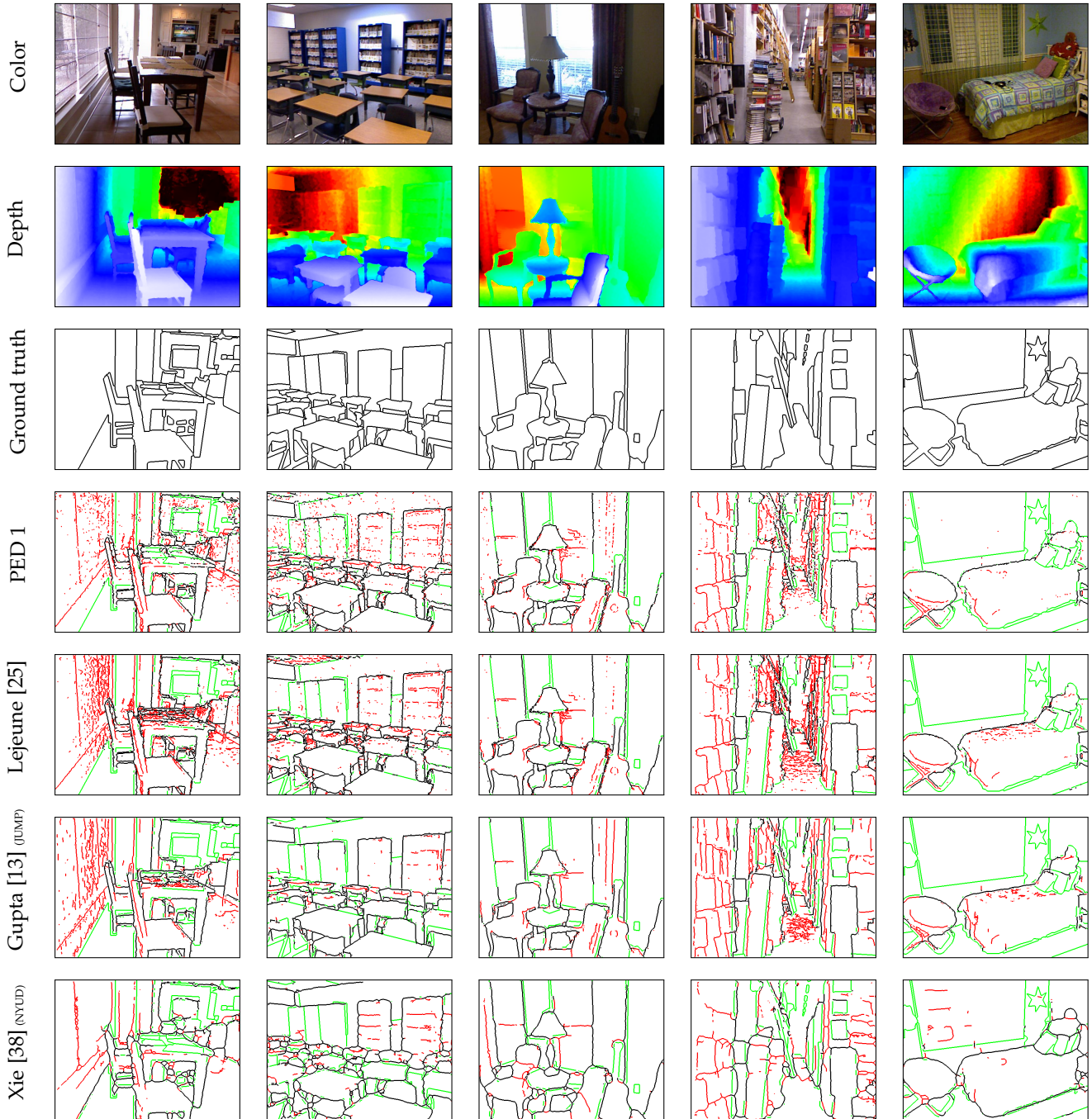


Figure 9. Edge detection results on the NYU Depth Dataset. The shown edge images were generated for the parameters maximizing the ODS score. Black, red, and green pixels represent the true positives, false positives, and false negatives, respectively.

APPENDIX A

CONDITIONAL PROBABILITY DENSITY FUNCTION OF RANGE VALUES ON SURFACE

A.1 Background and notations

Consider the configuration shown in Fig. 3. Our goal is to find a geometrical relationship between the exact ranges \bar{z}_p and \bar{z}_q and the signed angle θ between (1) the direction from \bar{Q} to \bar{P} and (2) the reference direction l_{om} related to the pixels p and q corresponding to \bar{P} and \bar{Q} .

To do so, it is convenient to use the normalized (signed)

range difference

$$\delta z_{pq} = \frac{\bar{z}_p - \bar{z}_q}{\bar{z}_p + \bar{z}_q}. \quad (21)$$

We also introduce the pixel $m = (p + q)/2$, which is midway between p and q in the image plane.

We express \bar{P} as the 3D vector $(x_p^{3D}, y_p^{3D}, \bar{z}_p)^T$, and p as the 2D vector $(x_p^{2D}, y_p^{2D})^T$, which we also express in homogeneous coordinates as the vector $\tilde{p} = (x_p^{2D}, y_p^{2D}, 1)^T$. Corresponding coordinates in the real world and in the image plane are related by $\bar{P} = \bar{z}_p K^{-1} \tilde{p}$, where K is the camera calibration matrix. If K is not provided by the

manufacturer, then one must calibrate the camera, *e.g.* using the technique of [14].

Next, we define the two vectors forming the angle θ . The first, $\mathbf{t}_{\mathbf{p}\mathbf{q}}$, characterizes the direction from $\bar{\mathbf{Q}}$ to $\bar{\mathbf{P}}$, and is defined as

$$\mathbf{t}_{\mathbf{p}\mathbf{q}} = \frac{\bar{\mathbf{P}} - \bar{\mathbf{Q}}}{\bar{z}_{\mathbf{p}} + \bar{z}_{\mathbf{q}}}. \quad (22)$$

By simple algebra, we can also express $\mathbf{t}_{\mathbf{p}\mathbf{q}}$ in terms of $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$,

$$\mathbf{t}_{\mathbf{p}\mathbf{q}} = K^{-1} \left(\delta z_{\mathbf{p}\mathbf{q}} \left(\frac{\tilde{\mathbf{p}} + \tilde{\mathbf{q}}}{2} \right) + \left(\frac{\tilde{\mathbf{p}} - \tilde{\mathbf{q}}}{2} \right) \right). \quad (23)$$

The second, \mathbf{l}_{om} , characterizes the direction from the optical center $\bar{\mathbf{O}}$ of the camera to \mathbf{m} in the image plane, and is defined as

$$\mathbf{l}_{\text{om}} = K^{-1} \tilde{\mathbf{m}} \quad \text{with} \quad \tilde{\mathbf{m}} = \frac{\tilde{\mathbf{p}} + \tilde{\mathbf{q}}}{2}. \quad (24)$$

It should be clear that \mathbf{l}_{om} is the point along the segment $\bar{\mathbf{O}}\mathbf{m}$ at unit range. It is not necessarily a unit vector. The vector $\mathbf{t}_{\mathbf{p}\mathbf{q}}$ can also be conveniently expressed using \mathbf{l}_{om} as $\mathbf{t}_{\mathbf{p}\mathbf{q}} = \delta z_{\mathbf{p}\mathbf{q}} \mathbf{l}_{\text{om}} + \mathbf{d}_{\mathbf{p}\mathbf{q}}$, where $\mathbf{d}_{\mathbf{p}\mathbf{q}} = K^{-1} (\tilde{\mathbf{p}} - \tilde{\mathbf{q}}) / 2$ is defined quite similarly to \mathbf{l}_{om} .

It should be clear that the useful range of θ is limited to $[0, \pi]$.

A.2 Geometrical relation between $\delta z_{\mathbf{p}\mathbf{q}}$ and θ

Since θ is the angle between \mathbf{l}_{om} and $\mathbf{t}_{\mathbf{p}\mathbf{q}}$, we have

$$\frac{\mathbf{t}_{\mathbf{p}\mathbf{q}} \circ \mathbf{l}_{\text{om}}}{\|\mathbf{t}_{\mathbf{p}\mathbf{q}}\| \|\mathbf{l}_{\text{om}}\|} = \cos \theta, \quad (25)$$

where \circ denotes the scalar product, and $\|\cdot\|$ the norm. Our goal is to turn Eq. (25) into an equation for $\delta z_{\mathbf{p}\mathbf{q}}$, and then to solve this equation for $\delta z_{\mathbf{p}\mathbf{q}}$, yielding an expression for $\delta z_{\mathbf{p}\mathbf{q}}$ in terms of θ and parameters that are fixed for a given pair of pixels \mathbf{p} and \mathbf{q} .

By squaring both sides of Eq. (25), we get, after some straightforward manipulations,

$$\delta z_{\mathbf{p}\mathbf{q}}^2 + 2 \frac{\mathbf{l}_{\text{om}} \circ \mathbf{d}_{\mathbf{p}\mathbf{q}}}{\|\mathbf{l}_{\text{om}}\|^2} \delta z_{\mathbf{p}\mathbf{q}} - \frac{\|\mathbf{d}_{\mathbf{p}\mathbf{q}}\|^2}{\|\mathbf{l}_{\text{om}}\|^2} \cot^2 \theta + \frac{(\mathbf{l}_{\text{om}} \circ \mathbf{d}_{\mathbf{p}\mathbf{q}})^2}{\|\mathbf{l}_{\text{om}}\|^4} \frac{1}{\sin^2 \theta} = 0. \quad (26)$$

This is obviously a second order equation for $\delta z_{\mathbf{p}\mathbf{q}}$. The solutions of this equation are found to be given by

$$\delta z_{\mathbf{p}\mathbf{q}} = a_{\mathbf{p}\mathbf{q}} \pm b_{\mathbf{p}\mathbf{q}} \cot \theta, \quad (27)$$

where

$$\begin{cases} a_{\mathbf{p}\mathbf{q}} = -\frac{\mathbf{l}_{\text{om}} \circ \mathbf{d}_{\mathbf{p}\mathbf{q}}}{\|\mathbf{l}_{\text{om}}\|^2} \\ b_{\mathbf{p}\mathbf{q}} = \sqrt{\frac{\|\mathbf{d}_{\mathbf{p}\mathbf{q}}\|^2}{\|\mathbf{l}_{\text{om}}\|^2} - a_{\mathbf{p}\mathbf{q}}^2} \end{cases} \quad (28)$$

We now examine whether both expressions for Eq. (27) are meaningful. Let us rewrite this equation as $\delta z_{\mathbf{p}\mathbf{q}} - a_{\mathbf{p}\mathbf{q}} = \pm b_{\mathbf{p}\mathbf{q}} \cot \theta$. After some straightforward algebra, we can express it as

$$\frac{\mathbf{t}_{\mathbf{p}\mathbf{q}} \circ \mathbf{l}_{\text{om}}}{\|\mathbf{l}_{\text{om}}\|^2} = \pm b_{\mathbf{p}\mathbf{q}} \cot \theta. \quad (29)$$

From Fig. 3, one sees that $\mathbf{t}_{\mathbf{p}\mathbf{q}} \circ \mathbf{l}_{\text{om}}$ and $\cot \theta$ always have the same sign. Therefore, since $b_{\mathbf{p}\mathbf{q}} \geq 0$, we must only keep the “+” sign in the last expression, and thus in Eq. (27), which thus reduces to

$$\delta z_{\mathbf{p}\mathbf{q}} = a_{\mathbf{p}\mathbf{q}} + b_{\mathbf{p}\mathbf{q}} \cot \theta. \quad (30)$$

A.3 Probability density function of $\Delta Z_{\mathbf{p}\mathbf{q}}$

At this point, we treat the above deterministic θ and $\delta z_{\mathbf{p}\mathbf{q}}$ as realizations of random variables (RVs) respectively denoted by Θ and $\Delta Z_{\mathbf{p}\mathbf{q}}$. The relation between these two RVs is given by Eq. (30), rewritten here as

$$\Delta Z_{\mathbf{p}\mathbf{q}} = a_{\mathbf{p}\mathbf{q}} + b_{\mathbf{p}\mathbf{q}} \cot \Theta. \quad (31)$$

This is usefully viewed as a change of RV, from Θ to $\Delta Z_{\mathbf{p}\mathbf{q}}$. The pdf for $\Delta Z_{\mathbf{p}\mathbf{q}}$ can thus be expressed in terms of the pdf for Θ via the standard expression [28]

$$f_{\Delta Z_{\mathbf{p}\mathbf{q}}}(\delta z_{\mathbf{p}\mathbf{q}}) = f_{\Theta}(\theta) \Big|_{\theta=g^{-1}(\delta z_{\mathbf{p}\mathbf{q}})} \left| \frac{d}{d\delta z_{\mathbf{p}\mathbf{q}}} g^{-1}(\delta z_{\mathbf{p}\mathbf{q}}) \right|, \quad (32)$$

where

$$\delta z_{\mathbf{p}\mathbf{q}} = g(\theta) = a_{\mathbf{p}\mathbf{q}} + b_{\mathbf{p}\mathbf{q}} \cot \theta, \quad \text{and} \quad (33)$$

$$\theta = g^{-1}(\delta z_{\mathbf{p}\mathbf{q}}) = \cot^{-1} \frac{\delta z_{\mathbf{p}\mathbf{q}} - a_{\mathbf{p}\mathbf{q}}}{b_{\mathbf{p}\mathbf{q}}}. \quad (34)$$

After some straightforward algebra, we get

$$f_{\Delta Z_{\mathbf{p}\mathbf{q}}}(\delta z_{\mathbf{p}\mathbf{q}}) = f_{\Theta} \left(\cot^{-1} \frac{\delta z_{\mathbf{p}\mathbf{q}} - a_{\mathbf{p}\mathbf{q}}}{b_{\mathbf{p}\mathbf{q}}} \right) \frac{b_{\mathbf{p}\mathbf{q}}}{\left[b_{\mathbf{p}\mathbf{q}}^2 + (\delta z_{\mathbf{p}\mathbf{q}} - a_{\mathbf{p}\mathbf{q}})^2 \right]}. \quad (35)$$

This result is valid for any pdf for Θ .

In the special case that is of interest here, *i.e.* where Θ is uniformly distributed between 0 and π , Eq. (35) reduces to

$$f_{\Delta Z_{\mathbf{p}\mathbf{q}}}(\delta z_{\mathbf{p}\mathbf{q}}) = \frac{1}{\pi} \frac{b_{\mathbf{p}\mathbf{q}}}{\left[b_{\mathbf{p}\mathbf{q}}^2 + (\delta z_{\mathbf{p}\mathbf{q}} - a_{\mathbf{p}\mathbf{q}})^2 \right]}. \quad (36)$$

The pdf $f_{\Delta Z_{\mathbf{p}\mathbf{q}}}(\delta z_{\mathbf{p}\mathbf{q}})$ is recognized as being a Cauchy pdf, also called Lorentzian pdf, denoted by $C(\delta z_{\mathbf{p}\mathbf{q}}; a_{\mathbf{p}\mathbf{q}}, b_{\mathbf{p}\mathbf{q}})$.

A.4 Probability density function of $\bar{Z}_{\mathbf{q}}$ for a fixed $\bar{z}_{\mathbf{p}}$

By solving Eq. (21) for $\bar{z}_{\mathbf{q}}$, we get

$$\bar{z}_{\mathbf{q}} = \frac{1 - \delta z_{\mathbf{p}\mathbf{q}}}{1 + \delta z_{\mathbf{p}\mathbf{q}}} \bar{z}_{\mathbf{p}}. \quad (37)$$

At this point, we treat the above deterministic $\bar{z}_{\mathbf{q}}$ and $\delta z_{\mathbf{p}\mathbf{q}}$ as realizations of RVs respectively denoted by $\bar{Z}_{\mathbf{q}}$ and $\Delta Z_{\mathbf{p}\mathbf{q}}$. Note that we continue to treat $\bar{z}_{\mathbf{p}}$ as a deterministic quantity (later, we will treat it as a realization of a RV $\bar{Z}_{\mathbf{p}}$). The relation between these two RVs is given by Eq. (37) rewritten here as

$$\bar{Z}_{\mathbf{q}} = \frac{1 - \Delta Z_{\mathbf{p}\mathbf{q}}}{1 + \Delta Z_{\mathbf{p}\mathbf{q}}} \bar{z}_{\mathbf{p}}. \quad (38)$$

This is usefully viewed as a change of RV, from $\Delta Z_{\mathbf{p}\mathbf{q}}$ to $\bar{Z}_{\mathbf{q}}$. The pdf for $\bar{Z}_{\mathbf{q}}$ can thus be expressed in terms of the pdf for $\Delta Z_{\mathbf{p}\mathbf{q}}$ via

$$f_{\bar{Z}_{\mathbf{q}}}(\bar{z}_{\mathbf{q}}) = f_{\Delta Z_{\mathbf{p}\mathbf{q}}}(\delta z_{\mathbf{p}\mathbf{q}}) \Big|_{\delta z_{\mathbf{p}\mathbf{q}}=g^{-1}(\bar{z}_{\mathbf{q}})} \left| \frac{\partial}{\partial \bar{z}_{\mathbf{q}}} g^{-1}(\bar{z}_{\mathbf{q}}) \right|, \quad (39)$$

where

$$\delta z_{\mathbf{p}\mathbf{q}} = g^{-1}(\bar{z}_{\mathbf{q}}) = \frac{\bar{z}_{\mathbf{p}} - \bar{z}_{\mathbf{q}}}{\bar{z}_{\mathbf{p}} + \bar{z}_{\mathbf{q}}}. \quad (40)$$

After some long, but straightforward algebra, we get

$$f_{\bar{Z}_{\mathbf{q}}}(\bar{z}_{\mathbf{q}}) = \frac{1}{\pi} \frac{\bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}}}{(\bar{z}_{\mathbf{q}} - \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}})^2 + (\bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}})^2}, \quad (41)$$

with

$$l_{\mathbf{p}\mathbf{q}} = \frac{(1 - a_{\mathbf{p}\mathbf{q}}^2 - b_{\mathbf{p}\mathbf{q}}^2)}{1 + a_{\mathbf{p}\mathbf{q}}^2 + b_{\mathbf{p}\mathbf{q}}^2 + 2a_{\mathbf{p}\mathbf{q}}} \quad (42)$$

$$s_{\mathbf{p}\mathbf{q}} = \frac{2b_{\mathbf{p}\mathbf{q}}}{1 + a_{\mathbf{p}\mathbf{q}}^2 + b_{\mathbf{p}\mathbf{q}}^2 + 2a_{\mathbf{p}\mathbf{q}}}. \quad (43)$$

The pdf $f_{\bar{Z}_{\mathbf{q}}}(\bar{z}_{\mathbf{q}})$ is recognized as being a Cauchy (or Lorentzian) pdf. It is interesting and remarkable that the change of RV (38) preserved the general nature of the pdf, i.e. Cauchy.

It is important to note that $l_{\mathbf{p}\mathbf{q}}$ and $s_{\mathbf{p}\mathbf{q}}$ only depend on the calibration matrix K of the camera, and on the location of the two pixels \mathbf{p} and \mathbf{q} .

A.5 Probability density function of $\bar{Z}_{\mathbf{q}}$ perturbed by additive Gaussian noise $N_{\mathbf{q}}$

Above, we were led to treat the range $\bar{z}_{\mathbf{q}}$ as a RV as a way to take into account the many possible orientations θ of the vector \vec{PQ} with respect to the reference orientation l_{om} corresponding to the two pixels \mathbf{p} and \mathbf{q} in the image plane.

At this point, we wish to consider the noise that affects the measurements of $z_{\mathbf{q}}$ by the camera. We model this noise as an additive Gaussian noise. We thus consider the new RV $Z_{\mathbf{q}} = \bar{Z}_{\mathbf{q}} + N_{\mathbf{q}}$, where $\bar{Z}_{\mathbf{q}}$ is the RV considered earlier, with pdf given by Eq. (41), and $N_{\mathbf{q}}$ is the Gaussian RV with a zero mean and a variance $\sigma^2 = \sigma_{\mathbf{q}}^2$, i.e. $\mathcal{N}(0, \sigma^2)$, with pdf

$$f_{N_{\mathbf{q}}}(n_{\mathbf{q}}) = G(n_{\mathbf{q}}; 0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{n_{\mathbf{q}}^2}{2\sigma^2}}. \quad (44)$$

The pdf of $Z_{\mathbf{q}}$ is thus the convolution of $f_{\bar{Z}_{\mathbf{q}}}(z_{\mathbf{q}})$ and $f_{N_{\mathbf{q}}}(z_{\mathbf{q}})$,

$$f_{Z_{\mathbf{q}}}(z_{\mathbf{q}}) = f_{\bar{Z}_{\mathbf{q}}}(z_{\mathbf{q}}) * f_{N_{\mathbf{q}}}(z_{\mathbf{q}}), \quad (45)$$

where $*$ denotes the convolution operation. From the domain of spectral line broadening in physics [16], it is well known that the convolution of a Cauchy profile and a Gaussian profile is a Voigt profile. For the Cauchy pdf $C(z_{\mathbf{q}}; \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}}, \bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}})$ of Eq. (36) and Gaussian pdf $G(z_{\mathbf{q}}; 0, \sigma)$ of Eq. (44), the corresponding Voigt pdf, denoted by $V(z_{\mathbf{q}}; \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}}, \bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}}, \sigma)$, is given by

$$V(z_{\mathbf{q}}; \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}}, \bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}}, \sigma) = C(z_{\mathbf{q}}; \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}}, \bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}}) * G(z_{\mathbf{q}}; \sigma). \quad (46)$$

Analytically, one generally expresses the Voigt pdf in terms of the Faddeeva function [1]. In practice, we use an approximation of V called the pseudo-Voigt function V_p , consisting of some linear combination of a Cauchy pdf and a Gaussian pdf

$$V_p(z_{\mathbf{q}}) = \eta C(z_{\mathbf{q}}; \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}}, \Gamma_V) + (1 - \eta) G(z_{\mathbf{q}}; \bar{z}_{\mathbf{p}} l_{\mathbf{p}\mathbf{q}}, \Gamma_V), \quad (47)$$

where Γ_V and η are functions of σ and $\bar{z}_{\mathbf{p}} s_{\mathbf{p}\mathbf{q}}$ [16]. Note that we are considering that $z_{\mathbf{q}}$ is affected by noise, but that $\bar{z}_{\mathbf{p}}$ is still fixed and exact. If we also want to consider the noise affecting the measured range $z_{\mathbf{p}}$, we must introduce a new Gaussian RV $Z_{\mathbf{p}} = \bar{z}_{\mathbf{p}} + N_{\mathbf{p}}$ where $N_{\mathbf{p}}$ is a Gaussian RV with zero mean and variance $\sigma_{\mathbf{p}}^2$ and, then, the RV $Z_{\mathbf{q}}$ is defined by

$$Z_{\mathbf{q}} = \frac{1 - \Delta Z_{\mathbf{p}\mathbf{q}}}{1 + \Delta Z_{\mathbf{p}\mathbf{q}}} (\bar{z}_{\mathbf{p}} + N_{\mathbf{p}}) + N_{\mathbf{q}}, \quad (48)$$

which would lead to a new and much more difficult problem since the first term on the right-hand side no longer follows a Cauchy pdf.

In practice, because $|\delta z_{\mathbf{p}\mathbf{q}}| \ll 1$, we use the approximation $Z_{\mathbf{q}} \approx [(1 - \Delta Z_{\mathbf{p}\mathbf{q}}) / (1 + \Delta Z_{\mathbf{p}\mathbf{q}})] z_{\mathbf{p}} + N_{\mathbf{p}} + N_{\mathbf{q}}$. If $N_{\mathbf{p}}$ and $N_{\mathbf{q}}$ are uncorrelated, then the variance σ^2 of $N_{\mathbf{p}} + N_{\mathbf{q}}$ is $\sigma_{\mathbf{p}}^2 + \sigma_{\mathbf{q}}^2$. Then, the pdf $f_{Z_{\mathbf{q}}}(z_{\mathbf{q}})$ that takes into account the noise on both $z_{\mathbf{p}}$ and $z_{\mathbf{q}}$ is also a Voigt pdf with $\sigma^2 = \sigma_{\mathbf{p}}^2 + \sigma_{\mathbf{q}}^2$. We have experimentally observed that this leads to an excellent approximation of $f_{Z_{\mathbf{q}}}(z_{\mathbf{q}})$.

REFERENCES

- [1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, 1964.
- [2] M. Adams. Amplitude modulated optical range data analysis in mobile robotics. In *IEEE Int. Conf. Robot. Autom. (ICRA)*, volume 2, pages 8–13, May 1993.
- [3] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011.
- [4] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1):231–243, Jan. 2004.
- [5] P. Boulanger, F. Blais, and P. Cohen. Detection of depth and orientation discontinuities in range images using mathematical morphology. In *IEEE Int. Conf. Pattern Recogn. (ICPR)*, volume 1, pages 729–732, June 1990.
- [6] A. Buch, J. Jessen, D. Kraft, T. Savarimuthu, and N. Krüger. Extended 3D line segments from RGB-D data for pose estimation. In *Scandinavian Conf. Image Anal. (SCIA)*, volume 7944, pages 54–65, Espoo, Finland, June 2013.
- [7] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, Nov. 1986.
- [8] S. Coleman, B. Scotney, and S. Suganthan. Edge detecting for range data using Laplacian operators. *IEEE Trans. Image Process.*, 19(11):2814–2824, Nov. 2010.
- [9] P. Dollar and L. Zitnick. Fast edge detection using structured forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(8):1558–1570, Aug. 2015.
- [10] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli. Depth mapping using projected patterns, 2010. US Patent Publication number 20100118123.
- [11] B. Günsel, A. Jain, and E. Panayirci. Reconstruction and boundary detection of range and intensity images using multiscale MRF representations. *Comp. Vision and Image Understanding*, 63(2):353–366, Mar. 1996.
- [12] S. Gupta, P. Arbeláez, and J. Malik. Perceptual organization and recognition of indoor scenes from RGB-D images. In *IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 564–571, Portland, Oregon, USA, June 2013.
- [13] S. Gupta, R. Girshick, P. Arbelaez, and J. Malik. Learning rich features from RGB-D images for object detection and segmentation. In *Eur. Conf. Comput. Vision (ECCV)*, Sept. 2014.
- [14] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.
- [15] J. Huang, A. Lee, and D. Mumford. Statistics of range images. In *IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, volume 1, pages 324–331, Hilton Head Island, SC, USA, June 2000.
- [16] T. Ida, M. Ando, and H. Toraya. Extended pseudo-Voigt function for approximating the Voigt profile. *Journal of Applied Crystallography*, 33(6):1311–1316, Dec. 2000.

- [17] X. Jiang and H. Bunke. Edge detection in range images based on scan line approximation. *Comp. Vision and Image Understanding*, 73(2):183–199, Feb. 1999.
- [18] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. A stereo machine for video-rate dense depth mapping and its new applications. In *IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 196–202, June 1996.
- [19] C. Kerl, M. Souiai, J. Sturm, and D. Cremers. Towards illumination-invariant 3D reconstruction using ToF RGB-D cameras. In *Int. Conf. 3D Vision (3DV)*, volume 1, pages 39–46, Tokyo, Japan, Dec. 2014.
- [20] K. Khoshelham. Accuracy analysis of Kinect depth data. *Int. Archives of the Photogrammetry, Remote Sens. and Spatial Inform. Sci. (ISPRS)*, XXXVIII-5/W12:133–138, 2011.
- [21] S. Konishi, A. Yuille, J. Coughlan, and S. Zhu. Statistical edge detection: Learning and evaluating edge cues. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(1):57–74, Jan. 2003.
- [22] R. Krishnapuram and S. Gupta. Morphological methods for detection and classification of edges in range images. *J. Math. Imaging and Vision*, 2(4):351–375, Nov. 1992.
- [23] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE J. Quantum Electron.*, 37(3):390–397, Mar. 2001.
- [24] A. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *Int. J. Comp. Vision*, 41(1-2):35–59, Jan. 2001.
- [25] A. Lejeune, S. Piérard, M. Van Droogenbroeck, and J. Verly. A new jump edge detection method for 3D cameras. In *IEEE Int. Conf. 3D Imaging (IC3D)*, pages 1–7, Liège, Belgium, Dec. 2011.
- [26] M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-of-Flight sensor calibration for accurate range sensing. *Comp. Vision and Image Understanding*, 114(12):1318–1328, Dec. 2010.
- [27] F. Mufti and R. Mahony. Statistical analysis of signal measurement in time-of-flight cameras. *ISPRS J. Photogrammetry and Remote Sens.*, 66(5):720–731, Sept. 2011.
- [28] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, 1991.
- [29] B. Parvin and G. Medioni. Adaptive multiscale feature extraction from range data. *Comp. Vision, Graph., and Image Process.*, 45(3):346–356, Mar. 1989.
- [30] H. Rapp, M. Frank, F. Hamprecht, and B. Jahne. A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time-of-flight-cameras. *Int. J. Intell. Syst. Techn. and Appl.*, 5(3/4):402–413, Nov. 2008.
- [31] X. Ren and L. Bo. Discriminatively trained sparse code gradients for contour detection. In *Adv. in Neural Inform. Process. Syst. (NIPS)*, pages 584–592, Dec. 2012.
- [32] H. Schafer, F. Lenzen, and C. Garbe. Depth and intensity based edge detection in time-of-flight images. In *Int. Conf. 3D Vision (3DV)*, pages 111–118, Seattle, WA, USA, June 2013.
- [33] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *IEEE Int. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 1297–1304, Providence, RI, USA, June 2011.
- [34] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. In *Eur. Conf. Comput. Vision (ECCV)*, volume 7576 of *Lecture Notes Comp. Sci.*, pages 746–760, Firenze, Italy, Oct. 2012.
- [35] B. Steder, R. Rusu, K. Konolige, and W. Burgard. Point feature extraction on 3D range scans taking into account object boundaries. In *IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 2601–2608, Shanghai, China, May 2011.
- [36] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *IEEE/RSJ Int. Conf. Intell. Robots and Syst. (IROS)*, pages 573–580, Vilamoura, Portugal, Oct. 2012.
- [37] T. Tang, W. Lui, and W. Li. A lightweight approach to 6-DOF plane-based egomotion estimation using inverse depth. In *Australasian Conf. Robot. and Autom. (ACRA)*, pages 1–10, Melbourne, Australia, Dec. 2011.
- [38] S. Xie and Z. Tu. Holistically-nested edge detection. In *IEEE Int. Conf. Comput. Vision (ICCV)*, pages 1395–1403, Santiago, Chile, Dec. 2015.
- [39] C. Ye and G. Hegde. Robust edge extraction for SwissRanger SR-3000 range images. In *IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 2437–2442, Kobe, Japan, May 2009.



Antoine Lejeune received a Master degree in engineering from the École Centrale Paris in 2009 and a Master degree in computer science and engineering from the University of Liège in 2010. Since then, he has been pursuing his doctoral degree at the University of Liège. His interests include image processing, computer vision, 3D vision, and machine learning.



Jacques G. Verly received the Ingénieur Électronicien degree from University of Liège, Belgium. Sponsored by the Belgian American Educational Foundation (BAEF), he attended Stanford University, where he received the M.S. and Ph.D. degrees in electrical engineering. From 1980 to 2000, he was at MIT Lincoln Laboratory, doing research in many areas, including signal processing, image processing, and computer vision for several imaging sensors (visible, IR, laser radar, SAR). Since 2000, he has been a

full professor in the Department of EECS of the University of Liège. He is a founder of the Laboratory for Signal and Image Exploitation (INTELSIG). He has been the Vice-Dean of the School of Engineering. His latest research areas are 3D/stereography, drowsiness monitoring, and nanosatellites. He is the instigator of the world leading events 3D Stereo MEDIA, now called Stereopsia and “World Immersive Forum”, and SomnoSafe. He is the initiator of the educational OUFIT-1 nanosatellite, the first 100% Belgian satellite. He is a co-founder of the Phasya company, specialized in drowsiness monitoring. In 2003-2004, he held one of the two prestigious Franquai Chairs at Université Libre de Bruxelles (ULB). He has over 360 publications and 2 US patents. He is a CRB Fellow of the BAEF.



Marc Van Droogenbroeck received the degree in Electrical Engineering and the Ph.D. degree from the University of Louvain (UCL, Belgium), in 1990 and 1994, respectively. While working toward the Ph.D. he spent two years with the Center of Mathematical Morphology (CMM), School of Mines of Paris, Fontainebleau, France. In April 1994, he joined the New Development Department of Belgacom. He was the Head of the Belgian Delegation within the ISO/MPEG Committee and served as a Representative to the World

Wide Web Consortium for two years. In 2003, he was a Visiting Scientist at CSIRO, Sydney, Australia. Since 1998, he has been a Member of the Faculty of Applied Sciences at the University of Liège, Belgium, where he is currently a Professor. His current interests include computer vision, 3D vision, machine learning, real time processing, motion analysis, gait analysis, telecommunications, positioning, and robotics.