**Noname manuscript No.**
(will be inserted by the editor)

# Image-Guided ToF Depth Upsampling: A Survey

**Iván Eichhardt · Dmitry Chetverikov · Zsolt Jankó**

Received: date / Accepted: date

**Abstract** Recently, there has been remarkable growth of interest in the development and applications of Time-of-Flight (ToF) depth cameras. However, despite the permanent improvement of their characteristics, the practical applicability of ToF cameras is still limited by low resolution and quality of depth measurements. This has motivated many researchers to combine ToF cameras with other sensors in order to enhance and upsample depth images. In this paper, we review the approaches that couple ToF depth images with high-resolution optical images. Other classes of upsampling methods are also briefly discussed. Finally, we provide an overview of performance evaluation tests presented in the related studies.

**Keywords** ToF cameras · depth images · optical images · depth upsampling · survey

## 1 Introduction

Image-based 3D reconstruction of static [111,121,49] and dynamic [125] objects and scenes is a core problem of computer vision. In the early years of computer vision, it was believed that visual information is sufficient for computer to solve the problem, as humans can perceive dynamic 3D scenes based on their vision. However, humans do not need to build precise 3D models of an environment to be able to act in the environment while numerous applications of computer vision require precise 3D reconstruction.

I. Eichhardt
Eötvös Loránd University and MTA SZTAKI, Budapest, Hungary

D. Chetverikov
Eötvös Loránd University and MTA SZTAKI, Budapest

Z. Jankó
MTA SZTAKI, Budapest

Today, different sensors and approaches are often combined to achieve the goal of building a detailed, geometrically correct and properly textured 3D or 4D (spatio-temporal) model of an object or a scene. Visual and non-visual sensor data are fused to cope with atmospheric haze [112], varying illumination, surface properties [56], motion and occlusion. This requires good calibration and registration of the modalities such as colour and infrared images, laser-measured data (LIDAR, hand-held scanners, Kinect), or ToF depth cameras. The output is typically a point cloud, a depth image, or a depth image with a colour value assigned to each pixel (RGBD).

A calibrated stereo rig is a widespread, classical device to acquire depth information based on **visual data** [111]. Since its baseline, i.e, the distance between the two cameras, is usually narrow, the resulting depth accuracy is limited. (By depth accuracy we mean the overall accuracy of depth measurement.) Wide-baseline stereo [121] can provide a better accuracy at the expense of more frequent occlusions and partial loss of spatial data. A collection of different-size, uncalibrated images of an object (or a video) can also be used for 3D reconstruction. However, this requires dense point correspondences (or dense feature tracking) across images/frames, which is not always possible.

Photometric stereo [49] applies a camera and several light sources to acquire the surface normals. The normal vectors are integrated to reconstruct the surface. The method provides fine surface details but suffers from less robust global geometry [92]. The latter is better captured by stereo methods that can be combined with photometric stereo [92] to obtain precise local and global geometry.

Shape acquisition systems using structured light [109, 26] contain one or two cameras and a projector that casts a specific, fixed or programmable, pattern onto the shape surface. Systems with programmable light pattern can achieve high precision of surface measurement.

The approaches to image-based 3D reconstruction listed above are the most widely used in practice. A number of other approaches to 'Shape-from-X' exist [124, 126], such as Shape-from-Texture, Shape-from-Shading, Shape-from-Focus, or Structure-from-Motion. These approaches are usually less precise and robust. They can be applied when high precision is not required, or as additional shape cues in combination with other methods. For example, Shape-from-Shading can be used to enhance fine shape details in RGBD data [144, 95, 44].

Among the **non-visual** sensors, the popular Kinect [148] can be used for real-time dense 3D reconstruction, tracking and interaction [57, 93]. The original device, Kinect I, combines a colour camera with a depth sensor projecting invisible structural light. In the Kinect II, the depth sensor is a ToF camera coupled with a colour camera. Currently, Kinect's resolution and precision are somewhat limited but still sufficient for applications in game industry and human-computer interaction (HCI). (See the study [94] for Kinect sensor noise analysis resulting in improved depth measurement.)

Different LIDAR devices [10, 38] have numerous applications in various areas including robot vision, autonomous vehicles, traffic monitoring, as well as scanning and 3D reconstruction of indoor and outdoor scenes, buildings and complete residential areas. They deliver point clouds with a measure of surface reflectivity assigned to each point.

Last but not least, ToF depth cameras [28, 113, 45] acquire low-resolution, registered depth and reflectance images at the rates suitable for real-time robot vision, navigation, obstacle avoidance, game industry and HCI.

This paper is devoted to a specific but critical aspect of ToF image processing, namely, to depth image upsampling. The upsampling can be performed in different ways. We give a survey of the methods that combine a low-resolution ToF depth image with a registered high-resolution optical image in order to refine the depth image resolution, typically by a factor of 4 to 16.

The rest of the paper is structured as follows. In Section 2, we discuss an important class of ToF cameras and compare their features to the features of three main image-based methods. Although our survey is devoted to image-guided depth upsampling, for the sake of completeness Section 3 gives a brief overview of upsampling with stereo and with multiple measurements, as well. Section 4 is a survey of depth upsampling based on a single optical image. In Section 5, we discuss the performance evaluation test results presented in the reviewed literature on depth upsampling. Finally, Section 6 provides further discussion, conclusion and outlook.

## 2 Time-of-Flight cameras

A recent survey [28] offers a comprehensive summary of the operation principles, advantages and limitations of ToF cameras. The survey [28] focuses on lock-in ToF cameras which are widely used in numerous applications, while the other category of ToF cameras, the pulse-based, is still rarely used. Our survey is also devoted to lock-in ToF cameras; for simplicity we will omit the term 'lock-in'.

ToF cameras [113, 102, 37] are small, compact, low-weight, low-consumption devices that emit infrared light and measure the time-of-flight to the observed object for calculating the distance to the object, usually called the depth. Contrary to LIDAR devices, ToF cameras have no mobile parts, and they capture depth images rather than point clouds. In addition to depth, ToF cameras deliver registered reflectance images of the same size and reliability values of depth measurements.

The main disadvantages of ToF cameras are their low resolution and significant acquisition noise. Although both resolution and quality are gradually improving, they are inherently limited by chip size and small active illumination energy, respectively. The highest currently available ToF camera resolution is QVGA ($320 \times 240$), with VGA ($640 \times 480$) being a target of future development. (See [89] for a systematic analysis of ground truth datasets and evaluation methods to assess the quality of ToF imaging data.)

Table 1 compares ToF cameras to three main image-based methods in terms of basic features. Stereo vision (SV) and structured light (SL) need to solve the correspondence, or matching, problem; the other two methods – photometric stereo (PS) and ToF – are correspondence-free. Of the four techniques, only ToF does not require extrinsic calibration. SV is a passive method, the rest use active illumination. This allows them to work with textureless surfaces when SV fails. On the other hand, distinct, strong textures facilitate the operation of SV but can deteriorate the performance of the active methods, especially when different textures cover the surface and its reflectance varies.

The active methods operate well in low lighting conditions when scene illumination is poor. Not surprisingly, passive stereo fails when visibility is low. The situation reverses for bright lighting that can prevent the operation of PS and reduce the performance of SL and ToF. In particular, bright lighting can increase ambient light noise in ToF [28] if ambient light contains the same wavelength as camera light. (A more recent report [75] claims that the bright lighting performance of ToF is good.) High-reflectivity surfaces can be a problem for all of the methods.

PS is efficient for neither outdoor nor dynamic scenes. SL can cope with time-varying surfaces, but currently it is not applied in outdoor conditions. Both SV and ToF can be used outdoor and applied to dynamic scenes, although the

**Table 1** Comparison of four techniques for depth measurement.

| | stereo vision | photometric stereo | structured light | ToF camera |
|---|---|---|---|---|
| correspondence | yes | no | yes | no |
| extrinsic calibration | yes | yes | yes | no |
| active illumination | no | yes | yes | yes |
| weak texture performance | weak | good | good | good |
| strong texture performance | good | medium | medium | medium |
| low light performance | weak | good | good | good |
| bright light performance | good | weak | medium/weak | medium |
| outdoor scene | yes | no | no | yes? |
| dynamic scene | yes | no | yes | yes |
| image resolution | camera dependent | camera dependent | camera dependent | low |
| depth accuracy | mm to cm | mm | $\mu$m to cm | mm to cm |

outdoor applicability of ToF cameras can be limited by their illumination energy and range [22,16], as well as by ambient light. Image resolution of the first three techniques depends on the camera and can be high, contrary to ToF cameras whose resolution is low. Depth accuracy of SV depends on the baseline and is comparable to that of ToF [75]. The other two techniques, especially SL, can yield higher accuracy.

From the comparison in Table 1, we observe that ToF cameras and passive stereo vision have complementary features. In particular, the influence of surface texture and illumination on the performance of the two techniques is just the opposite. As discussed in Section 4, this complementarity of ToF sensing and stereo has motivated researchers to combine the two sources of depth data in order to enhance applicability, accuracy and robustness of 3D vision systems.

ToF cameras have numerous applications. The related surveys [29,28] conclude that the most exploited feature of the cameras is their ability to operate without moving parts while providing depth maps at high frame rates. This capability greatly simplifies the solution of a critical task of 3D vision, the foreground-background separation. ToF cameras are exploited in robot vision [55] for navigation [135, 21,128,145] and 3D pose estimation and mapping [101,85, 34].

Further important application areas are 3D reconstruction of objects and environments [17,27,6,31,67,63], computer graphics [122,103,65] and 3DTV [120,118,133,134, 78]. (See study [116] for a recent survey of depth sensing for 3D television.) ToF cameras are applied in various tasks related to recognition and tracking of people [40,7,64] and parts of human body: hand [79,91], head [35] and face [91, 108]. Alenya et al. [1] use colour and ToF camera data to build 3D models of leaves for automated plant measurement. Additional applications are discussed in the recent book [37].

## 3 Upsampling with stereo and with multiple measurements

Low resolution and low signal-to-noise ratio are the two main disadvantages of ToF depth imagery. The goal of depth image upsampling is to increase the resolution and simultaneously improve image quality, in particular, near depth edges where surface discontinuities tend to result in erroneous or missing measurements [28]. In some applications, such as mixed reality, game industry and 3DTV, the depth edge areas are especially important because they determine occlusion and disocclusion of moving actors.

Approaches to depth upsampling can be categorised into three main classes [24]. In this survey, we discuss image-guided upsampling when a high-resolution optical image registered with a low-resolution depth image is used to refine the depth. However, for completeness we will now briefly discuss the other two classes, as well.

Note that most of the ToF depth upsampling methods surveyed in this paper deal with lateral depth enhancement. As already mentioned, some techniques for RGBD data processing [144,95,44] enhance fine shape details by calculating surface normals.

**ToF–stereo fusion** combines ToF depth with multicamera stereo data. A recent survey of this type of depth upsampling is available in [90]. Hansard et al. [45] discuss some variants of this approach and provide a comparative evaluation of several methods. The important issue of registering the ToF camera and the stereo data is also addressed. By mapping ToF depth values to the disparities of a high-resolution camera pair, it is possible to simultaneously upsample the depth values and improve the quality of the disparities [39]. Kim et al. [63] address the problem of sparsely textured surfaces and self-occlusions in stereo vision by fusing multicamera stereo data with multiview ToF sensor measurements. The method yields dense and detailed 3D models of scenes challenging for stereo alone while enhancing the ToF depth images. Zhu et al. [150,149,151] also explore the

complementary features of ToF cameras and stereo in order to improve accuracy and robustness.

Yang et al. [141] present a setup that combines a ToF depth camera with three stereo cameras and report on GPU-based, fast stereo depth frame grabbing and real-time ToF depth upsampling. The system fails in large surface regions of dark (e.g., black) colour that cause troubles to both stereo and ToF cameras. Bartczak and Koch [5] combine multiple high-resolution colour views with a ToF camera to obtain dense depths maps of a scene. Similar input data are used by Li et al. [73] who present a joint learning-based method exploiting differential features of the observed surface. Kang and Ho [60,51] report on a system that contains multiple depth and colour cameras.

Hahne and Alexa [41,42] claim that combination of ToF camera and stereo vision can provide enhanced depth data even without precise calibration. Kuhnert and Stommel [67] fuse ToF depth data with stereo data for real-time indoor 3D environment reconstruction in mobile robotics. Further methods are discussed in the recent survey [90]. A drawback of ToF–stereo is that it still inherits critical problems of passive stereo vision: the correspondence problem, the problem of textureless surfaces, and the problem of occlusions.

A natural way to improve resolution is to combine multiple measurements of an object. In optical imaging, numerous studies are devoted to super-resolution [131,129] or upsampling [23] of colour images. Fusing multiple ToF depth measurements into one image is sometimes referred to as **temporal and spatial upsampling** [24]. This type of depth upsampling is less widespread than ToF–stereo fusion and image-guided methods.

Hahne and Alexa [43] obtain enhanced depth images by adaptively combining several images taken with different exposure (integration) times. Their method is inspired by techniques applied in high dynamic range (HDR) imaging where different measures of image quality are used as weights for adaptive colour image fusion. For depth image fusion, the method [43] uses local measures of depth contrast, well-exposedness, surface smoothness, and uncertainty defined via signal entropy.

In [115,15], the authors acquire multiple depth images of a static scene from different viewpoints and merge them into a single depth map of higher resolution. An advantage of such approaches is that it does not need a sensor of another type. Working with depth images only allows one to avoid the so-called 'texture copying problem' of sensor fusion when contrast image textures tend to 'imprint' onto the upsampled depth image. This negative effect will be discussed later in relation to image-guided upsampling. A limitation of the methods [115,15] is that only static objects can be measured.

Mac Aodha et al. [83] use a training dataset of high-resolution depth images for patch-based upsampling of a low-resolution depth image. Although theoretically attractive, the method is too time-consuming for most applications. A somewhat similar patch-based approach is presented by Hornacek et al. [52] who exploit patch-wise self-similarity of a scene and search for patch correspondences within the input depth image. The method [52] aims at single image based upsampling while the algorithm [83] needs a large collection of high-resolution exemplars to search in. A drawback of the method [52] is that it relies on patch correspondences which may be difficult to obtain, especially for less characteristic surface regions.

Riegler et al. [104] use a deep network for single depth map super-resolution. The same problem is addressed in [3] using the Discrete Wavelet Transform and in [84] using subdictionaries of exemplars constructed from example depth maps. Finally, the patent [61] describes a method for combined depth filtering and resolution refinement.

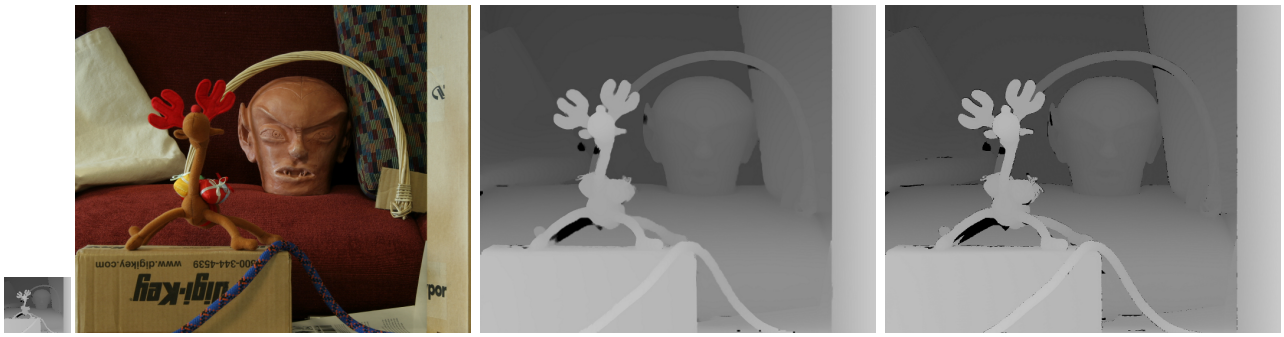
## 4 Image-guided depth upsampling

In this section, we provide a survey of depth upsampling based on a single optical image assuming calibrated and fused depth and colour data. As discussed later, precise calibration and sensor fusion are essential for good upsampling. Similarly to the ToF-stereo fusion survey [90], we classify the methods as local or global. The former category applies local filtering while the latter uses global optimisation. The approaches that fall in neither of the two classes are discussed separately.

We start the presentation of the methods by illustrating the upsampling problem, discussing its difficulties and introducing the notations. Fig. 1 shows an example of successful upsampling of a high-quality depth image of low resolution. The input depth and colour images are from the Middlebury stereo dataset [110]. The original high-resolution depth image was acquired with structured light, then artificially downsampled to get the low-resolution image shown in Fig. 1. Small parts of depth data (dark regions) are lost. The upsampled depth is smooth and very similar to the original high-resolution data used as the ground truth. In the Middlebury data, depth discontinuities match well the corresponding edges of the colour image. This dataset is often used for quantitative comparative evaluation of image-guided upsampling techniques.

For real-world data, the upsampling problem is more complicated than for the Middlebury data. Fig. 2 illustrates the negative features of depth images captured by ToF cameras[1]. The original depth resolution is very low compared to that of the colour image. When resized to the size of the colour image, the depth image clearly shows its deficiencies: a part of the data is lost due to low resolution; some

[1] Data courtesy of Zinemath Zrt [152].



input depth and colour images upsampled depth ground-truth depth

**Fig. 1** Sample Middlebury data, the upsampled depth and the ground truth.

shapes, e.g., the heads, are distorted. Despite the calibration, the contours of the depth image do not always coincide with those of the colour image. There are erroneous and missing measurements along the depth edges, in the dark region on the top, and in the background between the chair and the poster.

To use a high-resolution image for depth upsampling, one needs to relate image features to depth features. A basic assumption exploited by most upsampling methods is that image edges are related to depth edges, that is, to surfaces discontinuities. It is often assumed [18,33,81,97,98,74,24] that smooth depth regions exhibit themselves as smooth intensity/colour regions, while depth edges underlie intensity edges. We will call this condition the depth-intensity **edge coincidence assumption**.

Clearly, the assumption is violated in the regions of high-contrast texture and on the border of a strong shadow. Some studies [139,123] relax it in order to circumvent the problems discussed below and avoid the resulting artefacts. However, depth edges are in any case a sensitive issue. Since image features are the only data available for upsampling, one has to find a balance between the edge coincidence assumption and other priors. This balance is data-dependent, which may necessitate adaptive parameter tuning of an upsampling algorithm.

Precise camera calibration is crucial for the applications that require good-quality depth images, in general, and accurate depth discontinuities, in particular. Techniques and engineering tools used to calibrate ToF cameras and enhance their quality are discussed in numerous studies [77,50,45, 99,102,72,58]. Procedures for joint calibration of a ToF camera and an intensity camera are described in [97,98,24,132]. Many researchers apply the well-known calibration method [147]. A ToF camera calibration toolbox implementing the method presented in [69] is available at the web site [68].

Inaccurate registration of depth and intensity images due to imprecise calibration results in deterioration of the upsampled depth. Schwarz et al. [117] propose an error model for ToF sensor fusion and analyse relation between the model and inaccuracies in camera calibration and depth measurements. Xu et al. [137] address the problem of misalignment correction in the context of depth image-based rendering. Fig. 3 illustrates the effect of misalignment on depth upsampling. The discrepancy between the depth and intensity images is artificially introduced by a relative shift of two, five and ten pixels. As the shift grows, the depth borders become blurred and coarse.

Because of the optical radial distortion typical for many cameras, the discrepancy between the input images tends to grow with the distance from image centre. Fig. 4 shows an example of this phenomenon. The shape of the person in the centre of the scene in Fig. 4a is quite precise, with even fine details such as fingers being upsampled correctly. When the person moves to the periphery of the scene (Fig. 4b), his shape, e.g., in the region of the neck, becomes visibly distorted due to the growing misalignment.

Avoiding depth blur to preserve contrast depth edges is a major issue of upsampling methods. Because of the depth-intensity edge coincidence assumption, this issue is related to the texture copying (transfer) problem. Contrast image textures tend to exhibit themselves in the upsampled depth image as illustrated in Fig. 5 where textured regions cause visible perturbation in the refined depth. This disturbing phenomenon and possible remedies are discussed in the papers [139,123]. Further typical problems of image-guided depth upsampling are mentioned in Section 6.

In the sequel, we use the following notations:

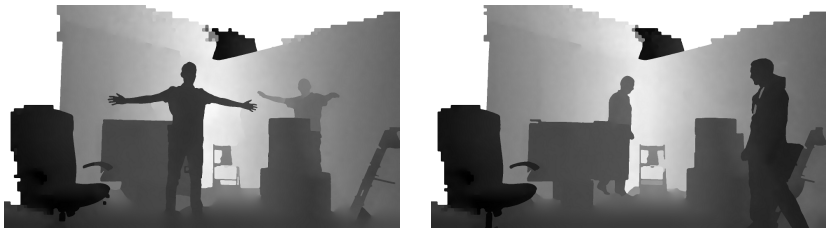
| | |
|---|---|
| $\mathtt{D}$ | Input (depth) image. |
| $\hat{\mathtt{D}}$ | Filtered / Upsampled image. |
| $\nabla\mathtt{D}$ | Gradient image. |
| $\tilde{\mathtt{I}}$ | Guide / Reference image. |
| $p,\ q, \ldots$ | 2D pixel coordinates. |
| $\|p-q\|$ | Distance between pixels $p$ and $q$. |
| $p_\downarrow,\ q_\downarrow, \ldots$ | Low-resolution coordinates, possibly fractional. |
| $\Omega(p)$ | A window around pixel $p$. |
| $\mathtt{D}_q$ | $\mathtt{D}$ value of pixel $q$. |

input depth and colour images | resized depth | upsampled depth

**Fig. 2** Data captured in a studio: the original depth and colour images, the depth image resized to colour image size, and the upsampled depth.

2 pixels | 5 pixels | 10 pixels

**Fig. 3** The effect of imprecise calibration on depth upsampling. The discrepancy between the input depth and colour images is 2, 5 and 10 pixels, respectively.



(a) (b)

**Fig. 4** The effect of optical radial distortion on depth upsampling.

$\|\mathtt{D}_p - \mathtt{D}_q\|$ Absolute difference of image values.
$f,\ g,\ h, \ldots$ Gaussian kernel functions.
$k_p$ Location-dependent normalisation factor: sum of weights in $\Omega(p)$.

## 4.1 Local methods

Image-guided ToF depth upsampling can be based on a single image or a video. Techniques using video rely on similar principles but they may exploit video redundancy and additional constraints such as motion coherence, also called temporal consistency. We will briefly discuss video-based approaches separately in Section 4.4.

The local methods use different forms of convolution with location-dependent weights $W\,(p,\,q)$:

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_q W\,(p,\,q)\,\mathtt{D}_q, \tag{1}$$

where

$$\sum_q \text{ stands for } \sum_{q \in \Omega(p)} \text{ and } k_p = \sum_q W\,(p,\,q)\,.$$

Upsampling techniques have to combine two different kinds of spatial data, ToF depth and intensity, or colour. When video is available, the temporal dimension should also be taken into account. Upsampling techniques based on filtering in spatial or spatio-temporal domain are often variants and extensions of the bilateral filter [130]. A bilateral

colour image upsampled depth ground truth

**Fig. 5** The texture transfer problem in depth upsampling.

filter $W_B(p, q)$ applies two Gaussian kernels, a spatial (or domain) one and a range one. The spatial kernel $g$ weights the distance from the filter center while the range kernel $f$ weights the absolute difference between the image value in the center and the value in a point of the window:

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_q W_B(p, q)\, \mathtt{D}_q, \tag{2}$$

where

$$W_B(p, q) = f(\|\mathtt{D}_p - \mathtt{D}_q\|)\, g(\|p - q\|). \tag{3}$$

The bilateral filter can be efficiently implemented in constant and real time [100, 140] which makes its practical application especially attractive. The reader is referred to the book [96] for a detailed discussion of bilateral filtering.

The idea of bilateral filtering has been extended in different ways. A joint (or cross) bilateral filter applies the range kernel to a second, guidance image $\tilde{\mathtt{I}}$ rather than to the input image $\mathtt{D}$:

$$W_{JB}(p, q) = f\left(\left\|\tilde{\mathtt{I}}_p - \tilde{\mathtt{I}}_q\right\|\right) g(\|p - q\|). \tag{4}$$

Note that $\mathtt{D}$ and $\tilde{\mathtt{I}}$ have the same resolution.

Joint bilateral filters have been successfully used in a wide range of tasks including the Joint Bilateral Upsampling (JBU) of depth images [66]. The input depth image $\mathtt{D}$ is assumed to be of lower resolution than the guidance image $\tilde{\mathtt{I}}$, thus the filter processes low-resolution pixel coordinates $q_\downarrow$. For values at fractional image coordinates, interpolation is assumed.

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_{q_\downarrow} W_{JBU}(p, q)\, \mathtt{D}_{q_\downarrow}, \tag{5}$$

where

$$W_{JBU}(p, q) = f\left(\left\|\tilde{\mathtt{I}}_p - \tilde{\mathtt{I}}_q\right\|\right) g(\|p_\downarrow - q_\downarrow\|). \tag{6}$$

Further attempts to combine different criteria and enhance the result of upsampling led to the use of multilateral [146, 80], rather than bilateral, filters. In particular, adding the median filter to the bilateral framework can improve the robustness of the method. The weighted median filter is defined as

$$\hat{\mathtt{D}}_p = \arg\min_b \sum_q W(p, q)\, |b - \mathtt{D}_q|. \tag{7}$$

The weighted median minimises the total weighted photometric distance from the central pixel to the other pixels of the window. (See [143] for a tutorial on weighted median filtering.) The Joint Bilateral Median (JBM) upsampling filter combines the median with $W_{JBU}$:

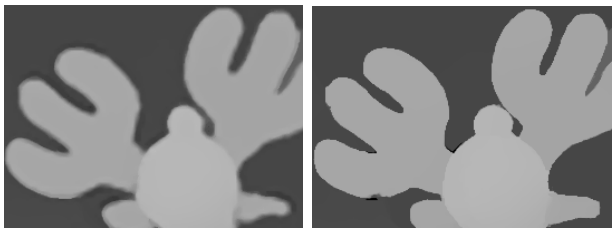
$$\hat{\mathtt{D}}_p = \arg\min_b \sum_{q_\downarrow} W_{JBU}(p, q)\, \left|b - \mathtt{D}_{q_\downarrow}\right|, \tag{8}$$

where $W_{JBU}$ is defined in (6).

Fig. 6 illustrates the difference between the Joint Bilateral and the Joint Bilateral Median upsampling filters. In both methods, bilateral weights are used. The main difference stems from the different effects of the weighted average of JB and the weighted median of JBM. While the former results in gradual blending and finer variation in depth, the latter allows for more drastic transitions and provides more contrast depth edges. The JB upsampling follows colour variations and is likely to result in depth interpolation. The JBM upsampling is more resistant to colour variations and outliers. This results in less depth interpolation and less texture transfer.

Chan et al. [12] propose an upsampling scheme based on the composite joint bilateral filter that locally adapts to the noise level and the smoothness of the depth function. The noise-aware filter [12] is defined as

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_{q_\downarrow} g(\|p_\downarrow - q_\downarrow\|) \tag{9}$$
$$\cdot \left[\alpha_p f_1\left(\left\|\tilde{\mathtt{I}}_p - \tilde{\mathtt{I}}_q\right\|\right) + (1 - \alpha_p)\, f_2\left(\left\|\mathtt{D}_{p_\downarrow} - \mathtt{D}_{q_\downarrow}\right\|\right)\right] \mathtt{D}_{q_\downarrow},$$



Joint Bilateral Joint Bilateral Median

**Fig. 6** Comparison of JB and JBM upsamplings.

where $f_1$ and $f_2$ are Gaussian kernels. Via the local context-sensitive parameter $\alpha_p$, the method blends the standard JBU ($\alpha_p = 1$) and an edge-preserving smoothing depth filter independent from colour data ($\alpha_p = 0$). Such solution can potentially reduce artefacts such as texture copying. Fu and Zhou [30] propose a combination of the noise-aware filter and a weighted mode filter with adaptive support window.

Riemens et al. [105] present a multi-step (multiresolution) implementation of JBU that doubles the depth resolution at each step. Garcia et al. [33] enhance the joint bilateral upsampling by taking into account the low reliability of depth values near depth edges. The pixel weighted average strategy [33] relies on the credibility map that depends on the depth gradient magnitude $\left\|\nabla \mathtt{D}_{q_\downarrow}\right\|$:

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_{q_\downarrow} h\left(\left\|\nabla \mathtt{D}_{q_\downarrow}\right\|\right) W_{JBU}\left(p,\, q\right) \mathtt{D}_{q_\downarrow}, \tag{10}$$

The credibility map $h\left(\left\|\nabla \mathtt{D}_{q_\downarrow}\right\|\right)$ prefers locations of moderate depth changes. (Recall $h$ is a Gaussian kernel.) The filter (10) tries to average over smooth surfaces while avoiding averaging across depth edges.

Yang et al. [142] apply the joint bilateral filter to a cost volume that measures the distance between potential depth candidates and the ToF depth image resized to the colour image size. The filter enforces the consistence of the cost values and the colour values. The upsampling problem is formulated as adaptive cost aggregation, a strategy frequently used in stereo matching [111,36]. To improve the robustness of the method [142] and its performance at depth edges, the authors add the weighted median filter and propose a multilateral framework [139]. The improved method [139] is implemented on a GPU to build a real-time high-resolution depth capturing system. Another cost-volume based technique using self-similarity matching is presented in the study [32].

The Non-Local Means (NLM) filter [9,2] can be viewed as a generalisation of the bilateral filter. In the photometric term of the bilateral similarity kernel, the bilateral filter uses point-wise intensity/colour difference while NLM uses patch-wise difference. Similarly, the geometric term of NLM relies on distance between patches rather than points. NLM allows for large (theoretically, infinite) distances resulting in strong contribution from distant patches. In this sense, NLM is theoretically a non-local filter. However, in practice the search for patches is limited to some neighbourhood, that is, the method is still more or less local. The photometric term assigns Gaussian weights to distant patch pixels, which gives greater importance to patch centres. See the recent survey [86] for a discussion of the NLM filter.

NLM has been successfully applied to depth upsampling [53] and enhancement [53,138]. The method proposed by Huhle et al. [53] applies the colour NLM filter including depth outlier detection and removal. The paper [53] discusses the interdependence between surface texturing and smoothing. The authors point out that the correspondence of depth and image pixels may change due to the displacement of the reconstructed point. Further cases of the application of NLM to depth upsampling will be discussed below in relation to global methods.

### 4.2 Global methods

The early paper [18] presents an application of Markov Random Field (MRF) to depth upsampling using a high-resolution colour image. The two-layer MRF is defined via the quadratic difference between measured and estimated depths, a depth smoothing prior, and the weighting factors that relate image edges to depth edges. This formulation leads to a least square optimisation problem which is solved by the conjugate gradient algorithm. Lu et al. [81] use a linear cost term (truncated absolute difference) since the quadratic cost is less robust to outliers. Their formulation of the MRF-based depth upsampling problem includes adaptive elements and is solved by the loopy belief propagation. Choi et al. [14] use quadratic terms in the proposed MRF energy and apply both discrete and continuous optimisation in a multiresolution framework.

A number of approaches [24,97,98] apply an optimisation algorithm to an upsampling cost function not related to an MRF. Such cost functions often contain terms similar to those used by the MRF-based methods. Ferstl et al. [24] define an energy functional that combines a standard quadratic depth data term with a regularising Total Generalised Variation [8] term and an anisotropic diffusion term that relates image gradients to depth gradients. As discussed in [4], anisotropic diffusion is closely related to bilateral filtering and adaptive smoothing. The primal-dual optimisation algorithm is used to minimise the energy functional. A MATLAB code of the upsampling approach [24], as well as synthetic and real benchmark data are available on the web site of the project [106].

Park et al. [97,98] apply an MRF to detect and remove outliers in depth data prior to upsampling. However, their

optimisation approach to upsampling does not rely on Markov Random Fields. The functional formulated in [97,98] includes a quadratic data term, a smoothness term and a Non-Local Means regularising term. The smoothness term combines segmentation, colour, edge saliency and depth information. The NLM regularising term is defined with the help of an anisotropic structure-aware filter. This term helps preserve local structure and fine details in presence of significant noise.

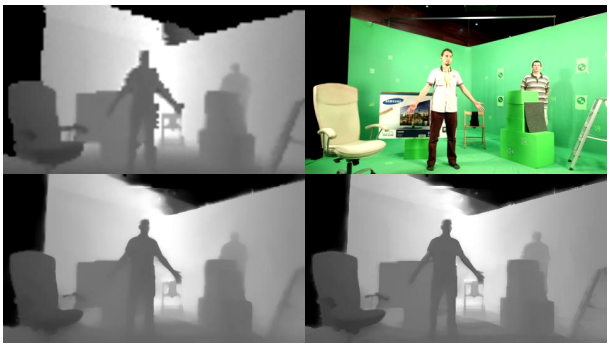
### 4.3 Other methods

Segmentation of colour and depth images can be used for upsampling either separately [127] or in combination with other tools. Tallon et al. [127] propose an upsampling and noise reduction method based on joint segmentation of depth and intensity into regions of homogeneous colour and depth. Conditional mode estimation is used to detect and correct regions with inconsistent features. Soh et al. [123] point out that the image-depth edge coincidence assumption may occasionally be invalid. They oversegment the colour image to obtain image super-pixels and use them for depth edge refinement. Then a maximum *a posteriori* probability [88] MRF framework is used to further enhance the depth.

Li et al. [74] develop a Bayesian approach to depth image upsampling that accounts for intrinsic camera errors. The method simulates uncertainty of depth and colour measurements by a Gaussian and a spatial-anisotropic kernel, respectively. The scene is assumed to be piecewise planar. The Random Sample Consensus (RANSAC) algorithm [25] is applied to select inliers for each plane model. An objective function combining depth and colour data terms is introduced and optimised to obtain the refined depth.

A promising research direction is the application of deep learning in order to avoid explicit filter construction and hand-designed objective functions. Li et al. [76] use a Convolutional Neural Network (CNN) to build a joint filter for depth upsampling. To enhance the depth image, Hui et al. [54] use a deep multi-scale convolutional network that learns high-resolution features in the optical image.

Most of the above mentioned studies compare the proposed method to existing techniques. Often, images from the Middlebury stereo dataset [110] containing the ground truth depth are used for quantitative comparison. The evaluation study by Langmann et al. [71] uses images from [110] as well as manually labelled ToF camera and colour data. The study compares a number of image-guided upsampling methods including bilateral filters, MRF optimisation [18] and the cost volume-based technique [142]. In Section 5 devoted to comparative evaluation studies, we will discuss the main conclusions of the paper [71].



first frame

second frame

**Fig. 7** Illustration of video-based depth upsampling. For each frame, the upper row shows the resized depth image and the corresponding optical image. The lower row shows upsampling results without (left) and with (right) temporal coherence.

### 4.4 Video-based depth upsampling

In this section, we briefly discuss the depth upsampling methods that use video rather than a single image. As already mentioned, the two categories of methods are based on the same assumptions and principles, but the video-based techniques may apply additional constraints. Fig. 7 illustrates the process of video-based upsampling. Two frames of a colour video sequence and a synchronised depth video sequence are demonstrated along with two different upsampling results. For the first result shown on the left-hand side, each frame was processed separately. (Compare to Fig. 2 where another single-image based upsampling algorithm was applied.) The method that yields the second result utilises temporal coherence with optical flow[2]. One can observe that the second result is, generally, better, except for a few locations such as the blurred contour of the person in the background.

To obtain depth video, Choi et al. [13] apply motion-compensated frame interpolation and the composite Joint Bilateral Upsampling procedure [12]. Dolson et al. [19] consider dynamic scenes and do not use the assumption of identical frame rate of the two video streams. They present a

[2] The methods have been developed by the authors of this survey. The algorithms are presented in [20].

Gaussian framework for multidimensional extension of 2D bilateral filter in space and time. A fast GPU implementation is discussed.

Xian et al. [136] consider synchronised depth and optical video cameras and propose upsampling solution implemented on a GPU in real time on the frame-by-frame basis without temporal processing. Their multilateral filter is inspired by the composite Joint Bilateral Upsampling procedure [12]. Kim et al. [62] propose a depth video upsampling method that also operates on the frame-by-frame basis. They use an adaptive bilateral filter taking into account the low signal-to-noise ratio of ToF camera data. The problem of texture copying is addressed.

Richardt et al. [103] consider the task of video-based upsampling in the context of computer graphics applications, such as video relighting, geometry-based abstraction and stylisation, and rendering. The depth data are first pre-processed to remove typical artefacts. Then a dual-joint bilateral filter is applied to upsample the depth. Finally, a spatio-temporal filter is used that blends the spatial and temporal components. A blending parameter specifies the degree of depth propagation from the previous to the current time steps using motion compensation.

Min et al. [87] propose a weighted mode filter based on a joint histogram. The temporal coherence of the depth video is achieved by extending the method to the neighboring frames. Optical flow supported by a patch-based flow reliability measure is used for motion estimation and compensation. In the studies [118–120], the authors view the depth upsampling process as a weighted energy optimisation problem constrained by temporal coherence. The space-time redundancy of intensity and depth is exploited in [59].

Vosters et al. [133] evaluate and compare several efficient video depth upsampling methods in terms of depth accuracy and interpolation quality in the context of 3DTV. They also present an analysis of computational complexity and runtime for GPU implementations of the methods. In a further study [134], the authors discuss 3DTV requirements for a high-quality depth map and propose a subsampling method based on the algorithms [87] and [33]. The study [134] also provides a benchmark and qualitative analysis of temporal post-processing methods in depth upsampling for 3DTV.

## 5 Comparative evaluation studies

We have already mentioned several studies that introduce novel methods for depth upsampling and compare them to a number of alternative techniques. In this section, we discuss these experimental performance evaluation results in more detail and summarise the conclusions of the comparative evaluations.

The survey of ToF-stereo fusion [90] has a section devoted to the evaluation of fusion methods. Different benchmark datasets and performance metrics are discussed. In relation to the Middlebury dataset [110], the authors criticise the often used approach when the original high-resolution ground truth depth is simply downsampled and some noise is added to the obtained depth map. Two additional aspects, sensor data alignment and ToF sensor simulation [89] are considered to generate more realistic synthetic ToF images. The authors provide a collection of datasets at their web site [48].

Hansard et al. [45] compare different variants of ToF-stereo fusion using the stereo algorithm [11] based on the seed growing principle. Two real and three synthetic datasets are used in the tests that evaluate the original method [11] with colour image seeds and fusion algorithms with ToF depth seeds and various cost functions combining image and depth likelihoods. It is demonstrated that depth-guided seed growing yields significantly better results than the original stereo algorithm.

Park et al. [97] compare their NLM filtering method for image-guided depth upsampling to several state-of-the-art techniques. Quantitative test results for noise-free and noisy synthetic data are provided. Three datasets based on the Middlebury benchmark [110] are used. The input low-resolution depth images are downsampled Middlebury images for four different downsampling factors.

For the noise-free synthetic data, the method [97] is compared to the MRF-based approach [18], the bilateral filtering with volume cost refinement [142], and the guided image filtering [47]. The method [97] yields the highest accuracy in all cases, although the difference between the best result and the second best one is often minor.

As discussed in the study [90], performance on ideal data is not really indicative of the practical applicability of a method. To test robustness to depth noise, Park et al. [97] add Gaussian noise to the input depth images. In this test, the noise-aware bilateral filtering [12] is also included. For the noisy data, the NLM filtering [97] outperforms the other four methods in seven of the twelve cases. However, the method [142] provides comparable results as its accuracy is always close to that of the NLM; in four cases, it is even better.

Ferstl et al. [24] present test results for both synthetic and real data. In the first test, the noisy synthetic data of Park et al. [97] is used. The authors demonstrate that their optimisation algorithm outperforms the five methods compared in [97] in terms of accuracy and speed.

In the second test, the authors use the three real-world datasets [106] they created. Here the ground truth is measured by a high-resolution structured light scanner while the upsampling factor is around $6.25$. For the real data, the method [24] compares favourably to the joint bilateral up-

sampling [66] and the guided image filtering [47]. The authors provide their benchmarking framework [106] to facilitate quantitative comparison of methods on real data.

As already mentioned, the evaluation study by Langmann et al. [71] uses the Middlebury data as well as ground truth depth data created manually. The authors conclude that results for the two kinds of data are, generally, consistent. The only exception is the MRF approach [18] that performs significantly better on the real data. However, this method is much slower than the other techniques compared in the study, e.g., the cost volume technique [142] and the joint bilateral filter [66]. In terms of depth accuracy, the overall performance of the joint bilateral upsampling is found to be the best.

Li et al. [74] compare their algorithm to the joint bilateral upsampling [66], the guided image filtering [47], and the NLM filtering [97]. Selected noise-free Middlebury data is used along with sample data from the RGBD Object Dataset [70]. The former contains objects with curved surfaces, the latter objects with planar or less curved surfaces. Despite the assumption of piecewise planar surfaces used by the method [74], the results of the quantitative evaluation indicate its superior performance in terms depth accuracy. However, the use of noise-free data for curved surfaces and the very low upsampling rate ($\times 2$) set in the tests make the claim of superior performance less convincing.

In their experiments, Yang et al. [139] compare the proposed joint bilateral median upsampling (JBMU) to the original joint bilateral upsampling approach [66] and its extensions [105,53]. To measure the quality of the upsampled depth images, they calculate the percentage of bad pixels. (A pixel is called bad if its disparity error exceeds 1.) 37 noise-free datasets from the Middlebury benchmark are used for performance evaluation. The upsampling rates of $\times 4$, $\times 16$, $\times 64$ are tested demonstrating certain improvement in depth quality compared to the alternative techniques. Also, it is shown that JBMU is less vulnerable to texture copying.

The experimental studies discussed above often use the Root Mean Squared Error (RMSE) as the measure of inaccuracy, i.e, the difference between the upsampled depth and the ground truth. In general, this approach is acceptable, but in some applications another measure of accuracy can be preferable. For example, RMSE accumulated due to texture transfer is usually small while errors resulting from depth edge blur can be unproportionally large because of large depth discontinuities. In applications sensitive to texture transfer but less sensitive to depth edge blur, one should consider using a different error metrics.

Most of the comparative evaluation tests either use synthetic data and ignore the problem of sensor data alignment or solve the problem manually. As discussed in Section 4, imprecise alignment can lead to significant upsampling errors. In practice, especially in video-based depth upsampling, a good automatic solution to the alignment problem is required.

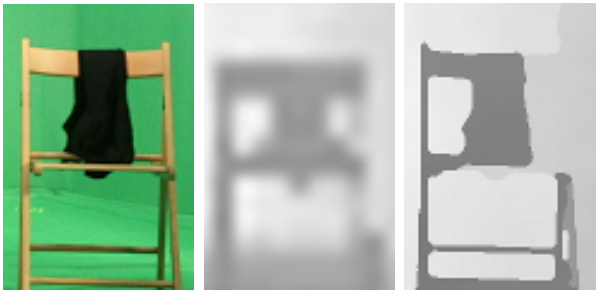



**Fig. 8** Illustration of the loss of narrow parts in upsampled depth data.

Local methods such as the filters discussed in Section 4.1 are usually faster and tend to respect fine details. Global consistency is not enforced explicitly, but it can be improved by multiscale iterative implementation of the filter. Global methods discussed in Section 4.2 provide better global consistency, often at the expense of higher computational cost and larger number of terms and parameters to tune. Some techniques such as cost aggregation [142,139] can be called 'semi-local' as they enhance global consistency by aggregation over support regions. This may involve increased memory usage and additional computational cost.

## 6 Discussion and conclusion

In Section 4, we mentioned some of the typical sources of errors in image-guided depth upsampling. In practice, one often faces further relevant problems such as the so-called 'flying pixels' at depth boundaries [72], flickering in video frames, occlusions due to disparity between the two cameras, and other sources of missing or outlying data, e.g., specular surfaces. Depth enhancement including completion of missing data [82] is addressed in numerous studies [134]. Below, we discuss two of these sources of errors that can result in loss or distortion of depth data.

Fig. 8 demonstrates an example of missing upsampled data for narrow parts of the chair in the background of the studio scene. (See Fig. 7.) Here, the low resolution of the depth camera prevents efficient operation of the upsampling algorithm despite the sufficient resolution of the optical image. When such narrow parts are essential but the depth camera resolution cannot be increased, one can resort to multiple measurements or detection and dedicated processing of the critical areas.

Fig. 9 illustrates the difficulty of processing shiny surfaces such as the metallic surface of a ladder. The quality of the upsampled depth is poor because the specific properties of the surface are not taken into account. While modelling

ladder upsampled depth

**Fig. 9** Illustration of poorly upsampled depth data for shiny surfaces.

of depth imaging systems [113] including analysis and modelling of their noise [94] has already been addressed, much less attention has been paid to surface-adaptive depth processing. We expect that future image-guided depth upsampling approaches will better adapt to scene context including geometry, reflectance properties, illumination, and motion.

The main purpose of this survey is to provide an introduction to the depth upsampling problem and give short descriptions of approaches. In our opinion, this problem is of interest beyond the area of ToF camera data processing since sensor data fusion becomes more and more popular. For example, studies in image-based point cloud upsampling [46, 114] apply tools similar or identical to those used by depth upsampling methods.

We believe that in near future ToF cameras will undergo fast changes in the direction of higher resolution, increasing range, better robustness and improved image quality. As a consequence, their application areas will extend and grow, leading to more frequent use and lower prices. (ToF camera in Kinect II is a definite step in this direction.) We also believe that the trend of coupling ToF cameras with other complementary sensors will persist resulting in growing demand for studies in depth data fusion with other kinds of data.

For the image processing community to be able to meet this demand, a critical issue is that of the evaluation and comparative testing of the proposed methods. Currently, many studies assume ideally calibrated data and provide tests on the Middlebury stereo dataset [110]. Such tests are not particularly indicative of performance in real applications. A good, rich benchmark of ToF data acquired in different real-world conditions is needed. The benchmark [106] providing datasets for three studio scenes is a step in this direction. The dataset [107] contains depth images and video sequences acquired by three different sensors. Other important related issues to be studied are sensor noise analysis [94] and sensor fusion error modelling and correction [117, 137].

## Acknowledgements

We are grateful to Zinemath Zrt for providing test data. This research was supported in part by the program “Highly industrialised region on the west part of Hungary with limited R&D capacity: Research and development programs related to strengthening the strategic future oriented industries manufacturing technologies and products of regional competences carried out in comprehensive collaboration” of the Hungarian National Research, Development and Innovation Fund (NKFIA), grant #VKSZ_12-1-2013-0038. This work was also supported by the NKFIA grant #K-120233.

## References

1. G. Alenya, B. Dellen, and C. Torras. 3D modelling of leaves from color and ToF data for robotized plant measuring. In *IEEE Int. Conf. on Robotics and Automation*, pages 3408–3414, 2011.
2. S.P. Awate and R.T. Whitaker. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 44–51, 2005.
3. C.S. Balure and M.R. Kini. Depth Image Super-Resolution: A Review and Wavelet Perspective. In *International Conference on Computer Vision and Image Processing*, pages 543–555, 2017.
4. D. Barash. Fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:844–847, 2002.
5. B. Bartczak and R. Koch. Dense depth maps from low resolution time-of-flight depth and high resolution color views. In *Advances in Visual Computing*, pages 228–239. Springer, 2009.
6. C. Beder, B. Bartczak, and R. Koch. A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
7. A. Bevilacqua, L. Di Stefano, and P. Azzari. People Tracking Using a Time-of-Flight Depth Sensor. In *Proc. Int. Conference on Video and Signal Based Surveillance*, page 89, 2006.
8. K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3:492–526, 2010.
9. A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, pages 490–530, 2005.
10. J. Carter, K. Schmid, K. Waters, L. Betzhold, B. Hadley, R. Mataosky, and J. Halleran. Lidar 101: An Introduction to Lidar Technology, Data, and Applications. Technical report, NOAA Coastal Services Center, Charleston, USA, 2012.
11. J. Čech and R. Šara. Efficient sampling of disparity space for fast and accurate matching. In *BenCOS Workshop, CVPR*, pages 1–8, 2007.
12. D. Chan, H. Buisman, C. Theobalt, and S. Thrun. A noise-aware filter for real-time depth upsampling. In *Proc. ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.
13. J. Choi, D. Min, B. Ham, and K. Sohn. Spatial and temporal up-conversion technique for depth video. In *Proc. Int. Conf. on Image Processing*, pages 3525–3528, 2009.
14. O. Choi, H. Lim, B. Kang, Y.S. Kim, K. Lee, J.D.K. Kim, and C.-Y. Kim. Discrete and continuous optimizations for depth image super-resolution. In *Proc. IS&T/SPIE Electronic Imaging*, pages 82900C–82900C, 2012.

15. Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 3D shape scanning with a time-of-flight camera. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1173–1180, 2010.
16. G. De Cubber, D. Doroftei, H. Sahli, and Y. Baudoin. Outdoor terrain traversability analysis for robot navigation using a time-of-flight camera. In *Proc. RGB-D Workshop on 3D Perception in Robotics*, 2011.
17. B. Dellen, R. Alenyà, Sergi Foix, S., and C. Torras. 3D object reconstruction from Swissranger sensor data using a spring-mass model. In *Proc. Int. Conf. on Comput. Vision Theory and Applications*, volume 2, pages 368–372, 2009.
18. J. Diebel and S. Thrun. An application of Markov random fields to range sensing. In *Proc. Advances in Neural Information Processing Systems*, pages 291–298, 2005.
19. J. Dolson, J. Baek, C. Plagemann, and S. Thrun. Upsampling range data in dynamic environments. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1141–1148, 2010.
20. I. Eichhardt, Z. Jankó, and D. Chetverikov. Novel methods for image-guided ToF depth upsampling. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 002073–002078, 2016.
21. P. Einramhof and M. Olufs, S. Vincze. Experimental evaluation of state of the art 3D-sensors for mobile robot navigation. In *Proc. Austrian Association for Pattern Recognition Workshop*, pages 153–160, 2007.
22. D. Falie and V. Buzuloiu. Wide range time of flight camera for outdoor surveillance. In *Proc. IEEE Symposium on Microwaves, Radar and Remote Sensing*, pages 79–82, 2008.
23. R. Fattal. Image upsampling via imposed edge statistics. *ACM Transactions on Graphics*, 26:95, 2007.
24. D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *Proc. Int. Conf. on Computer Vision*, pages 993–1000, 2013.
25. M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.
26. D. Fofi, T. Sliwa, and Y. Voisin. A comparative survey on invisible structured light. In *Electronic Imaging 2004*, pages 90–98, 2004.
27. S. Foix, G. Alenya, J. Andrade-Cetto, and C. Torras. Object modeling using a ToF camera under an uncertainty reduction approach. In *Proc. Int. Conf. on Robotics and Automation*, pages 1306–1312, 2010.
28. S. Foix, G. Alenya, and C. Torras. Lock-in Time-of-Flight (ToF) Cameras: A Survey. *Sensors Journal*, 11(9):1917–1926, 2011.
29. S. Foix, R. Alenyà, and C. Torras. Exploitation of time-of-flight (ToF) cameras. Technical Report IRI-DT-10-07, IRI-UPC, 2010.
30. M. Fu and W. Zhou. Depth map super-resolution via extended weighted mode filtering. In *Visual Communications and Image Processing*, pages 1–4, 2016.
31. S. Fuchs and S. May. Calibration and registration for precise surface reconstruction with time-of-flight cameras. *International Journal of Intelligent Systems Technologies and Applications*, 5:274–284, 2008.
32. N. Fukushima, K. Takeuchi, and A. Kojima. Self-similarity matching with predictive linear upsampling for depth map. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4, 2016.
33. F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier, and A. Cuesta. Pixel weighted average strategy for depth sensor data fusion. In *Proc. Int. Conf. on Image Processing*, pages 2805–2808, 2010.
34. P. Gemeiner, P. Jojic, and M. Vincze. Selecting good corners for structure and motion recovery using a time-of-flight camera. In *Int. Conf. on Intelligent Robots and Systems*, pages 5711–5716, 2009.
35. S.B. Gokturk and C. Tomasi. 3D head tracking based on recognition and interpolation using a time-of-flight depth sensor. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 211–217, 2004.
36. M. Gong, L. Yang, R.and Wang, and M. Gong. A performance study on different cost aggregation approaches used in real-time stereo matching. *International Journal of Computer Vision*, 75:283–296, 2007.
37. M. Grzegorzek, C. Theobalt, R. Koch, and A. Kolb (Eds). *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*. Springer, 2013.
38. H. Guan, J. Li, Y. Yu, M. Chapman, and C. Wang. Automated road information extraction from mobile laser scanning data. *Intelligent Transportation Systems, IEEE Transactions on*, 16:194–205, 2015.
39. S.A. Guomundsson, H. Aanæs, and R. Larsen. Fusion of stereo vision and time-of-flight imaging for improved 3D estimation. *Int. Journal of Intelligent Systems Technologies and Applications*, 5(3):425–433, 2008.
40. S.A. Guomundsson, R. Larsen, H. Aanæs, M. Pardas, and J.R. Casas. ToF imaging in smart room environments towards improved people tracking. In *Proc. Conf. on Computer Vision and Pattern Recognition Workshops*, pages 1–6, 2008.
41. U. Hahne and M. Alexa. Combining time-of-flight depth and stereo images without accurate extrinsic calibration. *Int. Journal of Intelligent Systems Technologies and Applications*, 5:325–333, 2008.
42. U. Hahne and M. Alexa. Depth imaging by combining time-of-flight and on-demand stereo. In *Dynamic 3D Imaging*, pages 70–83. Springer, 2009.
43. U. Hahne and M. Alexa. Exposure Fusion for Time-Of-Flight Imaging. In *Computer Graphics Forum*, volume 30, pages 1887–1894. Wiley Online Library, 2011.
44. Y. Han, J.-Y. Lee, and I. Kweon. High quality shape from a single RGD-D image under uncalibrated natural illumination. In *Proc. Int. Conf. on Computer Vision*, pages 1617–1624, 2013.
45. M. Hansard, S. Lee, O. Choi, and R. Horaud. *Time-of-flight cameras*. Springer, 2013.
46. A. Harrison and P. Newman. Image and sparse laser fusion for dense scene reconstruction. In *Field and Service Robotics*, pages 219–228. Springer, 2010.
47. K. He, J. Sun, and X. Tang. Guided image filtering. In *Proc. European Conf. on Computer Vision*, pages 1–14, 2010.
48. Heidelberg Collaboratory for Image Processing, Ruprecht-Karl University. Time of Flight Stereo Fusion Collection. `hci.iwr.uni-heidelberg.de/Benchmarks/`, 2016.
49. S. Herbort and C. Wöhler. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(3):1–17, 2011.
50. C. Herrera, J. Kannala, and J. Heikkilä. Joint depth and color camera calibration with distortion correction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34:2058–2064, 2012.
51. Y.-S. Ho and Y.-S. Kang. Multi-view depth generation using multi-depth camera system. In *International Conference on 3D Systems and Application*, pages 67–70, 2010.
52. M. Hornacek, C. Rhemann, M. Gelautz, and C. Rother. Depth Super Resolution by Rigid Body Self-Similarity in 3D. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1123–1130, 2013.
53. B. Huhle, T. Schairer, P. Jenke, and W. Straßer. Fusion of range and color images for denoising and resolution enhancement with a non-local filter. *Computer Vision and Image Understanding*, 114:1336–1345, 2010.
54. T.-W. Hui, C.C. Loy, and X. Tang. Depth Map Super-Resolution by Deep Multi-Scale Guidance. In *European Conference on Computer Vision*, pages 353–369, 2016.

55. S. Hussmann and T. Liepert. Robot vision system based on a 3D-ToF camera. In *Proc. Conf. on Instrumentation and Measurement Technology*, pages 1–5, 2007.
56. I. Ihrke, K.N. Kutulakos, H.P.A. Lensch, M. Magnor, and W. Heidrich. State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS 2008 State of the Art Reports*, 2008.
57. S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, S. Hodges, P. Kohli, J. Shotton, A.J. Davison, and A. Fitzgibbon. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proc. ACM Symp. on User Interface Software and Technology*, pages 559–568, 2011.
58. J. Jung, J.-Y. Lee, Y. Jeong, and I. Kweon. Time-of-flight sensor calibration for a color and depth camera pair. *PAMI*, 37:1501–1513, 2015.
59. U.S. Kamilov and P.T. Boufounos. Depth superresolution using motion adaptive regularization. In *IEEE International Conference on Multimedia & Expo Workshops*, pages 1–6, 2016.
60. Y.-S. Kang and Y.-S. Ho. High-quality multi-view depth generation using multiple color and depth cameras. In *IEEE Int. Conf. on Multimedia and Expo*, pages 1405–1410, 2010.
61. S. Katz, A. Adler, and G. Yahav. Combined depth filtering and super resolution. US Patent 8,660,362. `www.google.com/patents/US8660362`, 2014.
62. C. Kim, H. Yu, and G. Yang. Depth super resolution using bilateral filter. In *Proc. Int. Congress on Image and Signal Processing*, volume 2, pages 1067–1071, 2011.
63. Y.M. Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Miscusik, and S. Thrun. Multi-view image and ToF sensor fusion for dense 3D reconstruction. In *ICCV Workshops*, pages 1542–1549, 2009.
64. S. Knoop, S. Vacek, and R. Dillmann. Sensor fusion for 3D human body tracking with an articulated 3D body model. In *Proc. Int. Conf. on Robotics and Automation*, pages 1686–1691, 2006.
65. A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. In *Computer Graphics Forum*, volume 29, pages 141–159, 2010.
66. J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. In *ACM Transactions on Graphics*, volume 26, pages 673–678, 2007.
67. K.-D. Kuhnert and M. Stommel. Fusion of stereo-camera and pmd-camera data for real-time suited precise 3D environment reconstruction. In *Proc. Int. Conf. on Intelligent Robots and Systems*, pages 4780–4785, 2006.
68. A. Kuznetsova. A ToF camera calibration toolbox. `github.com/kapibara/ToF-Calibration`, 2015.
69. A. Kuznetsova and B. Rosenhahn. On calibration of a low-cost time-of-flight camera. In *ECCV Workshop on Consumer Depth Cameras for Computer Vision*, 2014.
70. K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view RGB-D object dataset. In *IEEE International Conference on Robotics and Automation*, pages 1817–1824, 2011.
71. B. Langmann, K. Hartmann, and O. Loffeld. Comparison of depth super-resolution methods for 2D/3D images. *Int. Journal of Computer Information Systems and Industrial Management Applications*, 3:635–645, 2011.
72. D. Lefloch, R. Nair, F. Lenzen, H. Schäfer, L. Streeter, M.J. Cree, R. Koch, and A. Kolb. Technical Foundation and Calibration Methods for Time-of-Flight Cameras. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 3–24. Springer, 2013.
73. J. Li, Z. Lu, G. Zeng, R. Gan, L. Wang, and H. Zha. A Joint Learning-Based Method for Multi-view Depth Map Super Resolution. In *Proc. Asian Conference on Pattern Recognition*, pages 456–460, 2013.
74. J. Li, G. Zeng, R. Gan, H. Zha, and L. Wang. A Bayesian approach to uncertainty-based depth map super resolution. In *Proc. Asian Conf. on Computer Vision*, pages 205–216, 2012.
75. Larry Li. Time-of-Flight Camera – An Introduction. Technical Report SLOA190B, Texas Instruments, 2014. Available at `www.ti.com/lit/wp/sloa190b/sloa190b.pdf`.
76. Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep Joint Image Filtering. In *European Conference on Computer Vision*, pages 154–169, 2016.
77. M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114:1318–1328, 2010.
78. M. Liu, Y. Zhao, J. Liang, C. Lin, H. Bai, and C. Yao. Depth Map Up-sampling with Fractal Dimension and Texture-Depth Boundary Consistencies. *Neurocomputing*, 2017.
79. X. Liu and K. Fujimura. Hand gesture recognition using depth data. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pages 529–534, 2004.
80. K.-H. Lo, Y.-C. Wang, and K.-L. Hua. Edge-Preserving Depth Map Upsampling by Joint Trilateral Filter. *IEEE Transactions on Cybernetics*, 2017.
81. J. Lu, D. Min, R.S. Pahwa, and M.N. Do. A revisit to MRF-based depth map super-resolution and enhancement. In *Int. Conference on Acoustics, Speech and Signal Processing*, pages 985–988, 2011.
82. S. Lu, X. Ren, and F. Liu. Depth Enhancement via Low-rank Matrix Completion. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 3390–3397, 2014.
83. O. Mac Aodha, N.D.F. Campbell, A. Nair, and G.J. Brostow. Patch based synthesis for single depth image super-resolution. In *Proc. European Conf. on Computer Vision*, pages 71–84, 2012.
84. S. Mandal, A. Bhavsar, and A.K. Sao. Depth Map Restoration From Undersampled Data. *IEEE Transactions on Image Processing*, 26:119–134, 2017.
85. S. May, D. Droeschel, D. Holz, C. Wiesen, and S. Fuchs. 3D pose estimation and mapping with time-of-flight cameras. In *Proc. IROS Workshop on 3D Mapping*, 2008.
86. Peyman Milanfar. A tour of modern image filtering: new insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine*, 30:106–128, 2013.
87. D. Min, J. Lu, and M.N. Do. Depth video enhancement based on weighted mode filtering. *IEEETIP*, 21:1176–1190, 2012.
88. Kevin P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
89. R. Nair, S. Meister, M. Lambers, M. Balda, H. Hofmann, A. Kolb, D. Kondermann, and B. Jähne. Ground Truth for Evaluating Time of Flight Imaging. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 52–74. Springer, 2013.
90. R. Nair, K. Ruhl, F. Lenzen, S. Meister, H. Schäfer, C.S. Garbe, M. Eisemann, M. Magnor, and D. Kondermann. A Survey on Time-of-Flight Stereo Fusion. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 105–127. Springer, 2013.
91. H. Nanda and K. Fujimura. Visual tracking using depth data. In *Proc. Conf. on Computer Vision and Pattern Recognition Workshops*, 2004.
92. D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. In *ACM Transactions on Graphics*, volume 24, pages 536–543, 2005.
93. R.A. Newcombe, A.J. Davison, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *Proc. IEEE Int. Symp. on Mixed and Augmented Reality*, pages 127–136, 2011.
94. C.V. Nguyen, S. Izadi, and D. Lovell. Modeling Kinect Sensor Noise for Improved 3D Reconstruction and Tracking. In *Second Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 524–530, 2012.

95. R. Or-El, G. Rosman, A. Wetzler, R. Kimmel, and A.M. Bruckstein. RGBD-Fusion: Real-Time High Precision Depth Recovery. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 5407–5416, 2015.
96. S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. *Bilateral Filtering*. Now Publishers Inc., 2009.
97. J. Park, H. Kim, Y.-W. Tai, M.S. Brown, and I. Kweon. High quality depth map upsampling for 3D-ToF cameras. In *Proc. Int. Conf. on Computer Vision*, pages 1623–1630, 2011.
98. J. Park, H. Kim, Y.-W. Tai, M.S. Brown, and I. Kweon. High-quality depth map upsampling and completion for RGB-D cameras. *IEEE Trans. Image Processing*, 23:5559–5572, 2014.
99. N. Pfeifer, D. Lichti, J. Böhm, and W. Karel. 3D cameras: Errors, calibration and orientation. In *TOF Range-Imaging Cameras*, pages 117–138. Springer, 2013.
100. F. Porikli. Constant time O(1) bilateral filtering. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
101. A. Prusak, O. Melnychuk, H. Roth, and I. Schiller. Pose estimation and map building with a time-of-flight-camera for robot navigation. *Int. Journal of Intelligent Systems Technologies and Applications*, 5:355–364, 2008.
102. F. Remondino and D. Stoppa. *ToF range-imaging cameras*. Springer, 2013.
103. C. Richardt, C. Stoll, N.A. Dodgson, H.-P. Seidel, and C. Theobalt. Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos. In *Computer Graphics Forum*, volume 31, pages 247–256, 2012.
104. G. Riegler, M. Rüther, and H. Bischof. ATGV-net: Accurate depth super-resolution. In *European Conference on Computer Vision*, pages 268–284, 2016.
105. A.K. Riemens, O.P. Gangwal, B. Barenbrug, and R.-P.M. Berretty. Multistep joint bilateral depth upsampling. In *IS&T/SPIE Electronic Imaging*, pages 72570M–72570M, 2009.
106. Robot Vision Laboratory, Graz University of Technology. ToFMark – Depth Upsampling Evaluation Dataset. `rvlab.icg.tugraz.at/tofmark/`, 2014.
107. D. Rotman and G. Gilboa. A depth restoration occlusionless temporal dataset. In *International Conference on 3D Vision*, pages 176–184, 2016.
108. J.R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez. Improving human face detection through ToF cameras for ambient intelligence applications. In *Ambient Intelligence-Software and Applications*, pages 125–132. Springer, 2011.
109. J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43:2666–2680, 2010.
110. D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. Middlebury stereo datasets. `vision.middlebury.edu/stereo/data/`, 2001–2014.
111. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.
112. L. Schaul, C. Fredembach, and S. Süsstrunk. Color Image Dehazing using the Near-Infrared. In *Proc. Int. Conf. on Image Processing*, pages 1629–1632, 2009.
113. Mirko Schmidt. *Analysis, Modeling and Dynamic Optimization of 3D Time-of-Flight Imaging Systems*. PhD thesis, Ruperto-Carola University of Heidelberg, Germany, 2011.
114. J.R. Schoenberg, A. Nathan, and M. Campbell. Segmentation of dense range information in complex urban scenes. In *Proc. Int. Conf. on Intelligent Robots and Systems*, pages 2033–2038. IEEE, 2010.
115. S. Schuon, C. Theobalt, J. Davis, and S. Thrun. Lidarboost: Depth superresolution for ToF 3D shape scanning. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 343–350, 2009.
116. S. Schwarz, R. Olsson, and M. Sjostrom. Depth Sensing for 3DTV: A Survey. *IEEE MultiMedia*, 20:10–17, 2013.
117. S. Schwarz, M. Sjöström, and R. Olsson. Multivariate Sensitivity Analysis of Time-of-Flight Sensor Fusion. *3D Research*, 5:1–16, 2014.
118. S. Schwarz, M. Sjöström, and R. Olsson. Temporal consistent depth map upscaling for 3DTV. In *IS&T/SPIE Electronic Imaging*, pages 901302–901302, 2014.
119. S. Schwarz, M. Sjöström, and R. Olsson. Weighted Optimization Approach to Time-of-Flight Sensor Fusion. *IEEE Trans. Image Processing*, 23:214–225, 2014.
120. Sebastian Schwarz. *Gaining Depth: Time-of-Flight Sensor Fusion for Three-Dimensional Video Content Creation*. PhD thesis, Mittuniversitetet, Sweden, 2014.
121. S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 519–528, 2006.
122. N. Snavely, C. L. Zitnick, S.B. Kang, and M. Cohen. Stylizing 2.5-D video. In *Proc. of 4th Int.l Symp. on Non-photorealistic Animation and Rendering*, pages 63–69. ACM, 2006.
123. Y. Soh, J.Y. Sim, C.S. Kim, and S.U. Lee. Superpixel-based depth image super-resolution. In *IS&T/SPIE Electronic Imaging*, pages 82900D–82900D. Int. Society for Optics and Photonics, 2012.
124. M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson, 2008.
125. E. Stoykova, A.A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, and S. Piekh. 3-D time-varying scene capture technologies – A survey. *IEEE Trans. on Circuits and Systems*, 17:1568–1586, 2007.
126. Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010.
127. M. Tallón, S.D. Babacan, J. Mateos, M.N. Do, R. Molina, and A.K. Katsaggelos. Upsampling and denoising of depth maps via joint-segmentation. In *Proc. of European Signal Processing Conference*, pages 245–249, 2012.
128. J.T. Thielemann, G.M. Breivik, and A. Berge. Pipeline landmark detection for autonomous robot navigation using time-of-flight imagery. In *Proc. Conf. on Computer Vision and Pattern Recognition Workshops*, pages 1–7, 2008.
129. J. Tian and K.-K. Ma. A survey on super-resolution imaging. *Signal, Image and Video Processing*, 5:329–342, 2011.
130. C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Int. Conf. on Computer Vision*, pages 839–846, 1998.
131. J.D. Van Ouwerkerk. Image super-resolution survey. *Image and Vision Computing*, 24:1039–1052, 2006.
132. V. Villena-Martínez, A. Fuster-Guilló, J. Azorín-López, M. Saval-Calvo, J. Mora-Pascual, J. Garcia-Rodriguez, and A. Garcia-Garcia. A Quantitative Comparison of Calibration Methods for RGB-D Sensors Using Different Technologies. *Sensors*, 17:243, 2017.
133. L. Vosters, C. Varekamp, and G. de Haan. Evaluation of efficient high quality depth upsampling methods for 3DTV. In *IS&T/SPIE Electronic Imaging*, pages 865005–865005, 2013.
134. L. Vosters, C. Varekamp, and G. de Haan. Overview of efficient high-quality state-of-the-art depth enhancement methods by thorough design space exploration. *Journal of Real-Time Image Processing*, pages 1–21, 2015.
135. J.W. Weingarten, G. Gruener, and R. Siegwart. A state-of-the-art 3D sensor for robot navigation. In *Proc. Int. Conf. on Intelligent Robots and Systems*, volume 3, pages 2155–2160, 2004.
136. X. Xiang, G. Li, J. Tong, M. Zhang, and Z. Pan. Real-time spatial and depth upsampling for range data. *Transactions on Computational Science XII: Special Issue on Cyberworlds*, 6670:78, 2011.

137. X. Xu, L.-M. Po, K.-H. Ng, L. Feng, K.-W. Cheung, C.-H. Cheung, and C.-W. Ting. Depth map misalignment correction and dilation for DIBR view synthesis. *Signal Processing: Image Communication*, 28:1023–1045, 2013.
138. K. Yang, Y. Dou, X. Chen, S. Lv, and P. Qiao. Depth Enhancement via Non-Local Means filter. In *International Conference on Advanced Computational Intelligence*, pages 126–130, 2015.
139. Q. Yang, N. Ahuja, R. Yang, K.-H. Tan, J. Davis, B. Culbertson, J. Apostolopoulos, and G. Wang. Fusion of Median and Bilateral Filtering for Range Image Upsampling. *IEEE Trans. Image Processing*, 22:4841–4852, 2013.
140. Q. Yang, K.-H. Tan, and N. Ahuja. Real-time O(1) bilateral filtering. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 557–564, 2009.
141. Q. Yang, K.H. Tan, B. Culbertson, and J. Apostolopoulos. Fusion of active and passive sensors for fast 3D capture. In *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, pages 69–74, 2010.
142. Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
143. L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. Weighted median filters: a tutorial. *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, 43:157–192, 1996.
144. L.-F. Yu, S.-K. Yeung, Y-W. Tai, and S. Lin. Shading-based shape refinement of RGBD images. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1415–1422, 2013.
145. F. Yuan, A. Swadzba, R. Philippsen, O. Engin, M. Hanheide, and S. Wachsmuth. Laser-based navigation enhanced with 3D time-of-flight data. In *Proc. Int. Conf. on Robotics and Automation*, pages 2844–2850, 2009.
146. L. Yuan, X. Jin, and C. Yuan. Enhanced Joint Trilateral Upsampling for Super-Resolution. In *Pacific Rim Conference on Multimedia*, pages 518–526, 2016.
147. Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:1330–1334, 2000.
148. Z. Zhang. Microsoft Kinect sensor and its effect. *IEEE MultiMedia*, 19:4–10, 2012.
149. J. Zhu, L. Wang, J. Gao, and R. Yang. Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32:899–909, 2010.
150. J. Zhu, L. Wang, R. Yang, and J.E. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
151. J. Zhu, L. Wang, R. Yang, J.E. Davis, and Z. Pan. Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(7):1400–1414, 2011.
152. Zinemath Zrt. The zLense platform. `www.zinemath.com/`, 2014.