2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Accurate Depth-Map Re nement by Per-Pixel Plane Fitting for Stereo Vision

Masashi Yokozuka, Kohji Tomita, Osamu Matsumoto and Atsuhiko Banno
National Institute of Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 Japan
Email: yokotsuka-masashi@aist.go.jp

***Abstract*T his paper discusses the re nement of sparse and noisy depth-maps to improve stereo measurements. Our method functions as a post- lter for stereo measurements, to remove outliers and interpolate the depths of invalid pixels. Per-pixel plane tting is employed to estimate the normals of an object s surface in a depth-map. These normals provide information regarding the interpolation of depth and the removal of outliers by evaluating the directions of surfaces. In our experiments, our method successfully reconstructed a dense and accurate geometry from a sparse and noisy depth-map, even where several dozen percent of pixels were outliers and only a few percent were from the original correct geometry. This result indicates a novel method of fast stereo measurement, because dense reconstruction can be performed without stereo matching for all pixels.**

I. Introduction

Visual perception of accurate scene geometry in real world applications remains a key problem in computer vision. In particular, it is challenging to achieve perfect scene reconstruction using a camera. In order to capture a complete and accurate geometrical picture using stereo vision, scene reconstruction must be able to determine the correct corresponding texture regions for all pixels from different views.

Because such complete stereo matching constitutes a challenging problem, we study depth-map re nement rather than improving stereo matching. Depth-map re nement aims to correct a depth-map after stereo measurement, by considering characteristics of the stereo measurement. This study discusses a method for the depth-map re nement of unreliable geometry obtained by a stereo measurement, as illustrated in Figure 1.

Depth-map re nement can be regarded as the denoising of a depth-map by employing an RGB-image, i.e., denoising for RGB-D images. There exists considerable research regarding denoising for RGB-D images obtained by one-shot capturing with an RGB-D camera [1] [5]. Such studies consider RGB-D cameras that can capture depth-maps almost correctly.

In this study, we consider the use of a standard camera, which has a much higher resolution and is usable in the

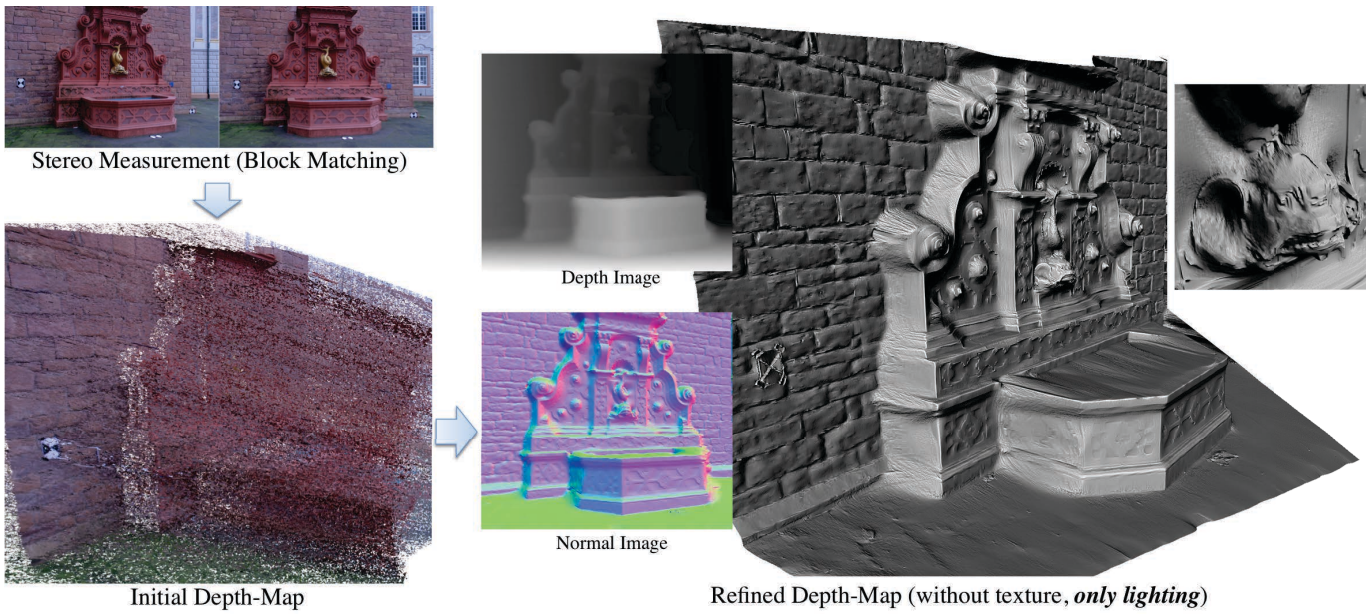

Fig. 1. Overview of our method for accurate depth-map re nement. Our method functions as a post- lter for stereo measurement. The method assumes an initial depth-map is given. This depth-map may include many outliers and invalid pixels, for which the depth cannot be measured. The proposed method removes outliers, and interpolates invalid pixels by per-pixel plane tting. By tting planes at every pixel in detail, the proposed method can reconstruct detailed shapes even if the input depth-map contains many outliers, as in this gure. The complicated central region is enlarged in the upper right gure to illustrate the details. We remark that this result was obtained from a pair of images, and multi-view stereo methods, such as in [9], were not employed.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

daytime. However, depth maps obtained by stereo measurement with a standard camera contain more noise and have many outliers and invalid pixels, for which depth cannot be measured.

The re nement of such depth-maps can be seen as a kind of global optimization, but it is not appropriate to apply standard global optimization techniques such as belief propagation [6] or graph-cut [7] and variational methods [8]. Global optimization techniques improve a depth-map as a combinatorial problem, involving nding a combination of depth values at all pixels satisfying minimum costs of the combined photometric and smoothness errors. To determine the optimal combination, global optimization techniques must store costs along all depth values at every pixel in the processing stage. For high resolution images, global optimization techniques face a memory problem in storing costs for all pixels.

To alleviate such problems, we propose a method that does not require global minimum costs to re ne depth values at every pixel. The idea is to consider characteristics of the disparity error distribution of a stereo measurement to reduce the computational load. Based on these characteristics, we propose a method of repeated per-pixel plane tting to re ne sparse and noisy depth-maps.

II. Characteristics of Stereo Measurements

First, we examine the characteristics of stereo measurements obtained from a preliminary experiment, in order to derive an approach to depth-map re nement.

A. Data set

For the investigation of characteristics, we use a data set created by Strecha *et al*. [9]. This data set was developed to evaluate methods for dense reconstruction using stereo vision, and contains ground-truth data obtained by laser-scanning.

B. Characteristics of block matching

Figure 2 illustrates the distribution of the disparity error obtained by block matching using normalized cross correlation (NCC). In this result, the disparity error distribution obtained by block matching is not a normal distribution, and is wide and thin except for a tall peak at zero error. This property implies that for correctly matched blocks, the precision of the depth falls within range of a few disparity errors. This result indicates that stereo measurement points are apt to be outliers when they fail in matching.

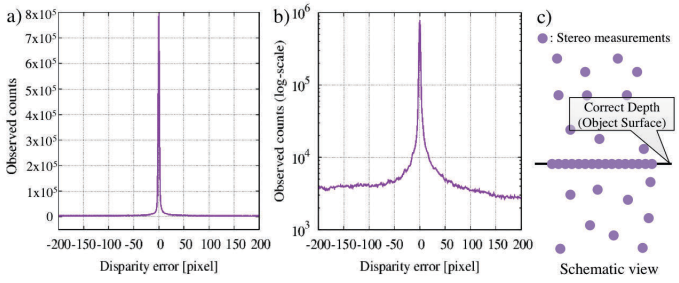

Fig. 2. Disparity error distribution obtained using block matching for the Initial Depth-Map in Figure 1. Histograms of observation counts according to the disparity error are shown in linear scale in a), and also in logarithmic scale in b) for clarity. Except for a tall peak at zero error, this distribution is wide and thin. Stereo measurements of the correct depth (object surface) under this distribution are schematically illustrated in c).

C. Relation between normals and RGB-images

In RGB-D images, color is strongly correlated with the normal of a plane. It follows that the possibility of two pixels in an RGB-D image lying on the same plane, i.e., the same object surface, is high when they have a similar color. In a previous study, in order to construct a cost function for stereo matching by considering surface directions, Bleyer *et al.* [10] evaluated whether two pixels lie on the same plane via the similarity between the two colors. As a similar concept, Yoon *et al.* [11] proposed an adaptive support weight that computes the likelihood of lying on the same plane through a bilateral-weight-like computation.

From these studies, we consider the following bilateral weight function:

$$W(x\ y\ \sigma_r\ \sigma_s) = W_r(x\ y\ \sigma_r)W_s(x\ y\ \sigma_s)$$
$$W_r(x\ y\ \sigma_r) = \frac{1}{N_r}\exp\left(-\frac{I_{xy} - I_{ij}\ ^2}{2\sigma_r^2}\right)$$
$$W_s(x\ y\ \sigma_s) = \frac{1}{N_s}\exp\left(-\frac{(x-i)^2 + (y-j)^2}{2\sigma_s^2}\right)$$

where $(i\ j)$ represents the coordinates of the center pixel, $(x\ y)$ describe the coordinates of a surrounding pixel, I is an RGB-image, $I_{ij}\ I_{xy}$ are RGB vectors at the corresponding coordinates, $N_r\ N_s$ are constants for normalization, and $\sigma_r\ \sigma_s$ are parameters for smoothness.[1] The weight function W_r computes the similarity between the center pixel and the surrounding pixels. The weight function W_s computes weights according to distance. This bilateral weight function computes the likelihood of whether two pixels lie on the same plane.

D. Approach to depth-map re nement

It follows from the characteristics of block matching that the precision of a depth measurement is very high when the matching is correct. When the matching is incorrect, the depth value has a high possibility of becoming an outlier. We assume that a depth-map can be re ned accurately if incorrect pixels can be removed, and following removal the invalid pixels can be interpolated using a plane reconstructed from correct pixels, because the correctly matched pixels have a highly precise depth.

Our depth-map re nement method involves the removal of outliers, with plane reconstruction per-pixel. Invalid pixels are interpolated by plane reconstruction, by employing bilateral weights that provide information regarding whether pixels lie

[1]The computational cost for a bilateral weight function is proportional to the value σ_s. For this reason, applying this function in image processing limits the window size for large σ_s. To solve this problem, many constant time bilateral lters have recently been proposed. In this study, we employ a domain transform lter [12] to apply large σ_s.

on corresponding planes. Outliers are detected by evaluating the distance from the reconstructed planes.

III. Depth-map Refinement including Invalid Pixels

A. Per-pixel plane estimation considering invalid pixels

First, we will describe the relation between a depth-map and per-pixel planes. The equation for a plane in three dimensions is described as $ax + by + cz + d = 0$. A three dimensional point $(x, y, z)^T$ is projected to a camera as $u = x/z$, $v = y/z$, and $\zeta = 1/z$, where (u, v) are normalized image coordinates and ζ is the inverse depth. If f_u, f_v represent the focal length and $(c_u, c_v)^T$ is a principal point, then the relation between the image coordinates (i, j) and normalized image coordinates (u, v) is described by $i = f_u u + c_u$ and $j = f_v v + c_v$. To obtain a relation between the normalized image coordinates (u, v), the inverse depth ζ, and a plane, we substitute (u, v) into $ax+by+cz+d = 0$. By simplifying the relation regarding the inverse depth ζ, this becomes

$$\zeta = \alpha u + \beta v + \gamma, \tag{1}$$

where $\alpha = -a/d$, $\beta = -b/d$, and $\gamma = -c/d$. The parameters α, β, γ characterize a normal of plane. When a unit normal $\mathbf{n} = (n_x, n_y, n_z)^T$ is employed, the plane equation becomes $n_x x + n_y y + n_z z + l = 0$, where $n_x^2 + n_y^2 + n_z^2 = 1$, $r = \sqrt{a^2 + b^2 + c^2}$, $n_x = a/r$, $n_y = b/r$, $n_z = c/r$, and $l = d/r$. By comparing the plane equation with α, β, γ, the following relation can be obtained for the normal of a plane:

$$n_x = -\alpha/\rho, \quad n_y = -\beta/\rho, \quad n_z = -\gamma/\rho, \tag{2}$$

where $\rho = \sqrt{\alpha^2 + \beta^2 + \gamma^2}$.

Before describing plane reconstruction, we will introduce some notation. When "g" is an image, we denote the value of the pixel at the image coordinates (i, j) as "g_{ij}." In addition, we denote the image domain as "Ω." Using these notations, we define the maps $u_{ij} = (i - c_u)/f_u$ and $v_{ij} = (j - c_v)/f_v$ of normalized image coordinates.

Next, we describe the estimation of a plane from a depth-map. The parameters α, β, γ can be estimated by linear regression, as described by (1). When points lie on an image domain Ω, the parameters α, β, γ are obtained by minimizing the following cost:

$$E = \sum_{i,j \in \Omega} \left(\zeta_{ij} - (\alpha u_{ij} + \beta v_{ij} + \gamma)\right)^2. \tag{3}$$

By minimizing (3), α, β, γ are given as

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \overline{u^2} & \overline{uv} \\ \overline{uv} & \overline{v^2} \end{pmatrix}^{-1} \begin{pmatrix} \overline{\zeta u} \\ \overline{\zeta v} \end{pmatrix}, \tag{4}$$
$$\gamma = \bar{\zeta} - \alpha \bar{u} - \beta \bar{v},$$

where the statistical values are

$$\begin{aligned} \overline{u} &= \frac{\sum_{i,j \in \Omega} u_{ij}}{|\Omega|}, & \overline{v} &= \frac{\sum_{i,j \in \Omega} v_{ij}}{|\Omega|}, \\ \overline{u^2} &= \frac{\sum_{i,j \in \Omega} u_{ij}^2}{|\Omega|}, & \overline{v^2} &= \frac{\sum_{i,j \in \Omega} v_{ij}^2}{|\Omega|}, \\ \overline{uv} &= \frac{\sum_{i,j \in \Omega} u_{ij} v_{ij}}{|\Omega|}, & \overline{\zeta} &= \frac{\sum_{i,j \in \Omega} \zeta_{ij}}{|\Omega|}, \\ \overline{\zeta u} &= \frac{\sum_{i,j \in \Omega} \zeta_{ij} u_{ij}}{|\Omega|}, & \overline{\zeta v} &= \frac{\sum_{i,j \in \Omega} \zeta_{ij} v_{ij}}{|\Omega|}, \end{aligned}$$

and $|\Omega|$ is the number of pixels in the domain Ω. This minimization result does not consider invalid pixels of an inverse depth-map ζ. To deal with invalid pixels, we employ a binary weight map w:

$$w_{ij} = \begin{cases} 1 & (\zeta_{ij} \text{ is a valid pixel}), \\ 0 & (\zeta_{ij} \text{ is an invalid pixel}). \end{cases} \tag{5}$$

Using the binary weight map w, the statistical values become

$$\begin{aligned} \overline{u} &= \frac{\sum_{i,j \in \Omega} w_{ij} u_{ij}}{\sum_{i,j \in \Omega} w_{ij}}, & \overline{v} &= \frac{\sum_{i,j \in \Omega} w_{ij} v_{ij}}{\sum_{i,j \in \Omega} w_{ij}}, \\ \overline{u^2} &= \frac{\sum_{i,j \in \Omega} w_{ij} u_{ij}^2}{\sum_{i,j \in \Omega} w_{ij}}, & \overline{v^2} &= \frac{\sum_{i,j \in \Omega} w_{ij} v_{ij}^2}{\sum_{i,j \in \Omega} w_{ij}}, \\ \overline{uv} &= \frac{\sum_{i,j \in \Omega} w_{ij} u_{ij} v_{ij}}{\sum_{i,j \in \Omega} w_{ij}}, & \overline{\zeta} &= \frac{\sum_{i,j \in \Omega} w_{ij} \zeta_{ij}}{\sum_{i,j \in \Omega} w_{ij}}, \\ \overline{\zeta u} &= \frac{\sum_{i,j \in \Omega} w_{ij} \zeta_{ij} u_{ij}}{\sum_{i,j \in \Omega} w_{ij}}, & \overline{\zeta v} &= \frac{\sum_{i,j \in \Omega} w_{ij} \zeta_{ij} v_{ij}}{\sum_{i,j \in \Omega} w_{ij}}. \end{aligned}$$

Finally, we describe per-pixel plane estimation for the reconstruction of depth and normals at each pixel. In order to expand the plane estimation at every pixel, our method employs convolution rather than summation:

$$f * g_{ij} = \iint_{\Omega} f(x, y) g(i - x, j - y) \mathrm{d}x \mathrm{d}y, \tag{6}$$

where f is a kernel (convolution matrix, or mask). When a Gaussian function is employed as a kernel, this convolution is equivalent to a Gaussian filter. Using convolution, the statistical values at coordinates (i, j) become

$$\begin{aligned} \overline{u_{ij}} &= \frac{f*(w_{ij} u_{ij})}{\overline{w}_{ij}}, & \overline{v_{ij}} &= \frac{f*(w_{ij} v_{ij})}{\overline{w}_{ij}}, \\ \overline{u_{ij}^2} &= \frac{f*(w_{ij} u_{ij}^2)}{\overline{w}_{ij}}, & \overline{v_{ij}^2} &= \frac{f*(w_{ij} v_{ij}^2)}{\overline{w}_{ij}}, \\ \overline{uv_{ij}} &= \frac{f*(w_{ij} u_{ij} v_{ij})}{\overline{w}_{ij}}, & \overline{\zeta_{ij}} &= \frac{f*(w_{ij} \zeta_{ij})}{\overline{w}_{ij}}, \\ \overline{\zeta u_{ij}} &= \frac{f*(w_{ij} \zeta_{ij} u_{ij})}{\overline{w}_{ij}}, & \overline{\zeta v_{ij}} &= \frac{f*(w_{ij} \zeta_{ij} v_{ij})}{\overline{w}_{ij}}, \end{aligned} \tag{7}$$

where $\overline{w}_{ij} = f * (w_{ij}) + \epsilon$ and ϵ is a small constant used to avoid division by zero. Because the results of convolution have a locality depending on the size of the kernel, the parameters α, β, γ are given by

$$\begin{pmatrix} \alpha_{ij} \\ \beta_{ij} \end{pmatrix} = \begin{pmatrix} \overline{u_{ij}^2} + \lambda & \overline{uv_{ij}} \\ \overline{uv_{ij}} & \overline{v_{ij}^2} + \lambda \end{pmatrix}^{-1} \begin{pmatrix} \overline{\zeta u_{ij}} \\ \overline{\zeta v_{ij}} \end{pmatrix}, \tag{8}$$
$$\gamma_{ij} = \overline{\zeta_{ij}} - \alpha_{ij} \overline{u_{ij}} - \beta_{ij} \overline{v_{ij}},$$

where λ is a stabilizing parameter for this linear equation. To obtain smoothed results, our method applies a smoothing filter to these results:

$$\overline{\alpha_{ij}} = f * \alpha_{ij}, \ \overline{\beta_{ij}} = f * \beta_{ij}, \ \overline{\gamma_{ij}} = f * \gamma_{ij}. \tag{9}$$

Finally, we obtain an inverse depth-map and a normal map by per-pixel plane estimation:

$$\begin{aligned} \widehat{\zeta}_{ij} &= \overline{\alpha_{ij}} u_{ij} + \overline{\beta_{ij}} v_{ij} + \overline{\gamma_{ij}}, \\ \mathbf{n}_{ij} &= -\frac{(\overline{\alpha_{ij}}, \overline{\beta_{ij}}, \overline{\gamma_{ij}})^T}{\sqrt{\overline{\alpha_{ij}}^2 + \overline{\beta_{ij}}^2 + \overline{\gamma_{ij}}^2}}, \end{aligned} \tag{10}$$

where $\widehat{\zeta}$ is an inverse depth-map including the interpolation of invalid pixels, and $\mathbf{n}$ is a normal map derived using plane estimation.

B. Implicit segmentation for guiding depth for invalid pixels

In order to guide depth and normals for invalid pixels, our method reconstructs a plane for the interpolation of invalid pixels. To reconstruct such a plane, it is necessary to gather points supporting the plane. In a depth-map, the gathering of points corresponds to the segmentation of the map. If this segmentation is correct, with each segment belonging to one plane, then the plane reconstruction becomes accurate.

In the place of segmentation, our method employs joint filtering. In the previous section, we demonstrated that a plane can be reconstructed by applying any filter to a depth-map. Joint filtering computes the weights of its kernel by considering the similarities of the color, intensity, or texture of a guide image. For example, the result of joint filtering applied to a depth-map is the weighted average of the depth values of neighboring pixels with similar colors. Thus, the statistical values in a region with similar pixels can be computed by joint filtering. Nevertheless, the region is not clearly segmented. For this reason, we refer to this statistical computation as implicit-segmentation.

We employ a joint bilateral filter, which is defined using bilateral weights as

$$f(I,\sigma_r,\sigma_s) * g_{ij} = \iint_{\Omega} W(x,y,\sigma_r,\sigma_s) g(i-x,j-y)\mathrm{d}x\mathrm{d}y.$$

C. Outlier removal with incremental thresholding

Our method removes outliers from an input depth-map by incremental thresholding against the distance from the reconstructed planes following per-pixel plane estimation. Outliers are removed by updating a binary weight map w:

$$w_{ij} = \begin{cases} 1 & (|z_{ij} - \widehat{z}_{ij}| \leq \theta\ \sigma_{ij} \cos\phi_{ij}), \\ 0 & (|z_{ij} - \widehat{z}_{ij}| > \theta\ \sigma_{ij} \cos\phi_{ij}), \end{cases} \tag{11}$$

where

$$\cos\phi_{ij} = -\mathbf{n}_{ij} \cdot \frac{(u_{ij}, v_{ij}, 1)^T}{\sqrt{u_{ij}^2 + v_{ij}^2 + 1}}.$$

Here, $z_{ij}(= 1/\zeta_{ij})$ is the depth from an input depth-map, $\widehat{z}_{ij}(= 1/\widehat{\zeta}_{ij})$ is the depth from a corrected inverse depth-map by per-pixel plane estimation, θ is the control variable for incremental thresholding, and σ_{ij} represents the uncertainness of the depth z_{ij}. The uncertainness σ_{ij} is derived from the depth uncertainty of a stereo measurement by discretization of the image domain. Furthermore, when the inverse depth ζ_{ij} is invalid, the uncertainness σ_{ij} is set to 0, for consistency of the algorithm. The control variable θ is updated at each iteration by

$$\theta \leftarrow \tau\theta, \tag{12}$$

while the range of the value of θ is $1 \leq \theta$. The parameter τ is set as $0 < \tau < 1$. Our method removes outliers by alternately repeating per-pixel plane estimation and thresholding, while decreasing the variable θ.

D. Summary of algorithm

A summary of our method is presented in **Algorithm 1**.

Algorithm 1 Per-pixel Plane Fitting

Input: ζ, w, σ, I, σ_r, σ_s, θ, τ, ϵ, λ
- ζ is an incomplete inverse depth-map.
- w is a binary weight map.
- σ is a map of depth uncertainness.
- I is a color (or monochrome) image.
- $\sigma_r, \sigma_s, \theta$, τ, ϵ, λ are parameters.

Output: $\widehat{\zeta}$, $\mathbf{n}$
- $\widehat{\zeta}$ is a corrected inverse depth-map.
- $\mathbf{n}$ is a normal map.

procedure PERPIXELPLANEFITTING(ζ, w, σ, I)
  while $\theta > 1$ **do**
    Compute $\overline{u}, \overline{v}, \overline{u^2}, \overline{v^2}, \overline{uv}, \overline{\zeta}, \overline{\zeta u}, \overline{\zeta v}$ with $f(I, \sigma_r, \sigma_s)$ and ϵ by (7).
    Compute α, β, γ with λ by (8).
    Compute $\overline{\alpha}, \overline{\beta}, \overline{\gamma}$ with $f(I, \sigma_r, \sigma_s)$ by (9).
    Compute $\widehat{\zeta}$,$\mathbf{n}$ by (10).
    Update w with θ and σ by (11).
    $\theta \leftarrow \tau\theta$
  end while
  return $\widehat{\zeta}$,$\mathbf{n}$
end procedure

IV. Experiments and Discussion

This paper discusses the refinement of depth-maps generated by stereo vision. The performance of a depth-map refinement is evaluated in terms of improvements in the accuracy.

A. Experimental method

First, we conduct a *performance evaluation experiment* to evaluate two aspects of the performance: noise-tolerance and interpolation. We generate an initial depth-map from ground-truth data as follows. For the noise-tolerance performance, we introduce noises to the depth-map, considering the characteristics of stereo measurement described in II-B. For the interpolation performance, we remove pixels from the generated depth-map. Second, we conduct an experiment regarding the *depth-map refinement from stereo measurement* using block-matching.

B. Evaluation criteria

The performance of our method is evaluated in terms of *completeness*, indicating the rate of pixels with correct depth values, defined by

$$\text{completeness} = \frac{\sum_{i,j\in\Omega} C_{ij}}{|\Omega|}, \tag{13}$$

where C_{ij} is the pixel-wise correctness, which is defined by

$$C_{ij} = \begin{cases} 1 & (|\widehat{z}_{ij} - Z_{ij}| < \Delta_1), \\ 0 & (|\widehat{z}_{ij} - Z_{ij}| \geq \Delta_1), \end{cases} \tag{14}$$

where $\widehat{z}_{ij}$ is the estimated depth, Z_{ij} is the true value of depth, and Δ_1 is the uncertainness of the stereo measurement, which indicates the range of depth with a 1 pixel disparity.

C. Experimental conditions

1) Data set: Similarly to the preliminary experiment, this experiment employs a data set created by Strecha *et al.* [9]. The considered images are views 6 and 7 in the *Fountain-P11* scene in the data set.

2) Stereo algorithm: In the second experiment, we employed a simple stereo algorithm, which is a block matching method using NCC. Moreover, we did not employ the commonly-used techniques of sub-pixel refinement and a left-right consistency check.

3) Parameters: The block size of the NCC was 9×9, and the image size was 3072×2048. The parameters for our method are $\theta = 30.0$, $\tau = 0.975$, $\sigma_r = 25/255$, $\sigma_s = 1024$, $\epsilon = 10^{-10}$, and $\lambda = 10^{-6}$. The parameter θ is an initial value. These parameters were obtained experimentally.[2]

D. Performance evaluation experiment

Our first experiment investigates the refinement performance. The following procedure is repeated by varying the rate of outliers and density of valid pixels:

1) Generate a depth-map from ground-truth data, with the addition of noise of a depth corresponding to a one pixel disparity using a normal distribution.
2) Replace pixels with a depth generated by a uniform distribution at a constant rate to randomly generate outliers.
3) Randomly remove pixels at a constant rate to create invalid pixels.
4) Apply our method to refine the depth-map.
5) Evaluate *completeness* of the refined depth-map against outliers and sparseness.

To describe the results, this paper uses the following values:

$$\text{Outlier rate} = \frac{\text{Num. of outliers}}{\text{Num. of all pixels}},$$

$$\text{Sampled density} = \frac{\text{Num. of sampled pixels}}{\text{Num. of all pixels}}.$$

Figure 3 presents examples of results for this experiment. Input geometries and reconstructed shapes are illustrated in this figure. Figure 4 presents the results of the performance evaluation for various sampled densities and outlier rates. It is a remarkable result that our method refined a shape with around 80% completeness when the outlier rate was 0% and sampled density was 0.5%. Even if the outlier rate of an input initial depth-map is over 50%, our method can reconstruct a dense geometry when the density of the geometry is over a few percent.

E. Depth-map refinement from stereo measurement

Next, we evaluate the recovering performance of our method for a depth-map obtained from a stereo measurement. The following procedure is repeated by varying the rates of outliers and the density of valid pixels.

[2]For large σ_s, the other parameters were not sensitive. The computational time required for the refinement of a depth-map was approximately 10 minutes using a Core-i7 PC.

1) Generate a depth-map by applying NCC based block matching.
2) Remove pixels by threshold processing of NCC.
3) Remove pixels randomly at a constant rate.
4) Apply our method to refine the depth-map.
5) Evaluate the *completeness* of the recovered depth-map against outliers and density.

Figure 5 presents performance curves. Before applying our method, this experiment performed threshold processing of NCC to alter the outlier rate. Fig. 1 presents the result obtained using all stereo measurements (the sampled density is 100%) without an NCC threshold (threshold is 0.0). These performance curves exhibit a similar trend as the performance curves from the first experiment, which used depth-maps generated from ground-truth data with added noise. Moreover, our method is able to reconstruct a dense and accurate geometry with 80% completeness from density of a few percent, even if depth-maps generated from a stereo measurement are used. This indicates a strong possibility of a novel stereo measurement method involving generating an accurate depth-map without stereo matching for all pixels. Because stereo matching involves many outliers and invalid pixels (e.g., pixels removed by NCC thresholding), our depth-map refinement method is suitable for stereo measurements, because it allows the inclusion of outliers and invalid pixels.

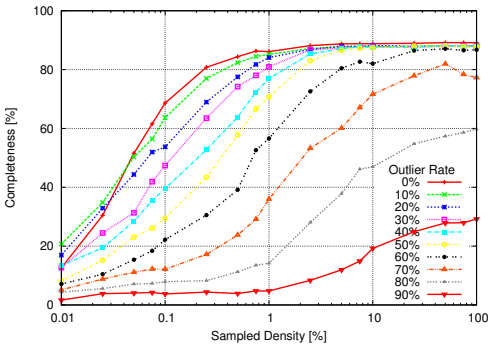

Fig. 4. Resulting curves of the *performance evaluation experiment* with varying outlier rates and sampled densities.

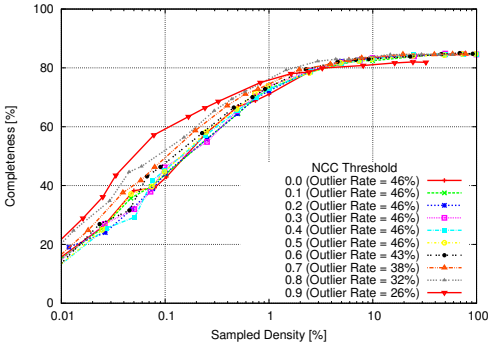

Fig. 5. Performance curves for the *depth-map refinement from a stereo measurement* with varying NCC thresholds and sampled densities.

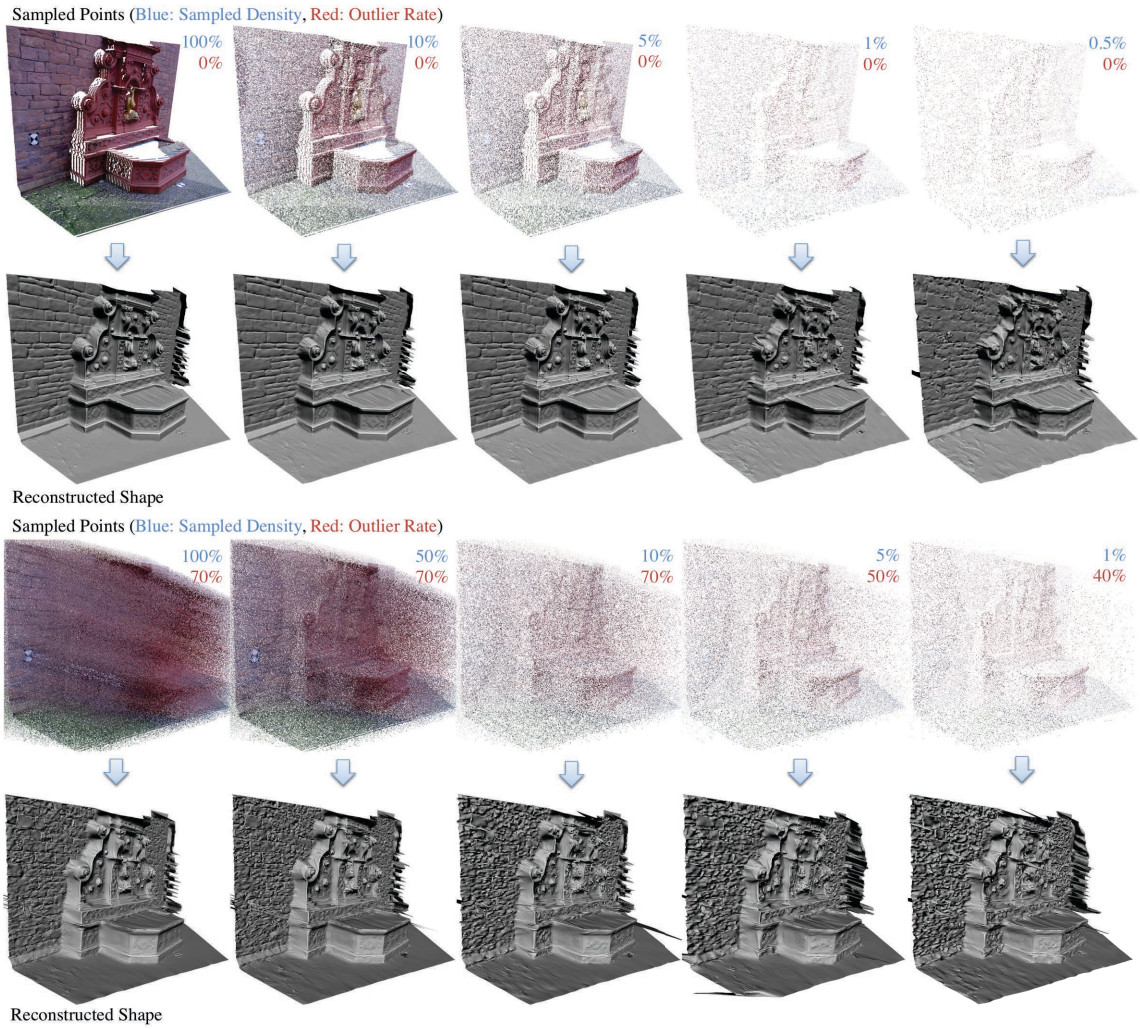

Fig. 3. Re nement examples for depth-maps generated from ground-truth data by adding noise and removing pixels.

V. Conclusion

This paper has proposed a depth-map re nement method that is suitable for stereo vision. Our method re nes a depth-map by per-pixel plane- tting, to remove outliers by evaluating the distance of a measurement point from a plane and considering the characteristics of stereo measurements. Our method successfully reconstructed a dense and accurate geometry from a noisy depth-map, in which the outlier rate was over 50%. A future study will include the evaluation of the reconstruction performance for many other depth-maps.

References

[1] K. Matsuo et al., Depth Image Enhancement Using Local Tangent Plane Approximations , in *Proc. of the Computer Vision and Pattern Recognition (CVPR)*, 2015.

[2] S. Lu et al., Depth enhancement via low-rank matrix completion, in *Proc. of the Computer Vision and Pattern Recognition (CVPR)*, 2014.

[3] M. Y. Liu et al., Joint Geodesic Upsampling of Depth Images, in *Proc. of the Computer Vision and Pattern Recognition (CVPR)*, 2013.

[4] M. Kiechle et al., A Joint Intensity and Depth Co-Sparse Analysis Model for Depth Map Super-Resolution, in *Proc. of the International Conference on Computer Vision (ICCV)*, 2013.

[5] L. Chen et al., Depth Image Enhancement for Kinect Using Region Growing and Bilateral Filter, in *Proc. of the International Conference on Pattern Recognition (ICPR)*, 2012.

[6] J. Sun et al., Stereo matching using belief propagation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787 800, 2003.

[7] V. Kolmogorov et al., Computing visual correspondence with occlusions using graph cuts, in *Proc. of the International Conference on Computer Vision (ICCV)*, 2001.

[8] A. Chambolle, An algorithm for total variation minimization and applications, *J. Mathematical Imaging and Vision*, vol.20, pages 88 97, 2004.

[9] C. Strecha et al., On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery, in *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2008.

[10] M. Bleyer et al., PatchMatch Stereo - Stereo Matching with Slanted Support Windows, in *Proc. of the British Machine Vision Conference (BMVC)*, 2011.

[11] A. J. Yoon et al., Locally adaptive support-weight approach for visual correspondence search, in *Proc. of the Computer Vision and Pattern Recognition (CVPR)*, 2005.

[12] E. Gastal and M. Oliveira, Domain Transform for Edge-Aware Image and Video Processing , in *Proc. of SIGGRAPH*, 2011.