

Dynamic 3D Scene Depth Reconstruction via Optical Flow Field Rectification

You Yang^{1,2,3}, Qiong Liu^{1,3}, Rongrong Ji^{3*}, Yue Gao^{4*}

1 Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China, **2** Department of Automation, Tsinghua University, Beijing, China, **3** Department of Electrical Engineering, Columbia University, New York, New York, United States of America, **4** School of Computing, National University of Singapore, Singapore, Singapore

Abstract

In this paper, we propose a depth propagation scheme based on optical flow field rectification towards more accurate depth reconstruction. In depth reconstruction, the occlusions and low-textural regions easily result in optical flow field errors, which lead ambiguous depth value or holes without depth in the obtained depth map. In this work, a scheme is proposed to improve the precision of depth propagation and the quality of depth reconstruction for dynamic scene. The proposed scheme first adaptively detects the occlusive or low-textural regions, and the obtained vectors in optical flow field are rectified properly. Subsequently, we process the occluded and ambiguous vectors for more precise depth propagation. We further leverage the boundary enhancement filters as a post-processing to sharpen the object boundaries in obtained depth maps for better quality. Quantitative evaluations show that the proposed scheme can reconstruct depth map with higher accuracy and better quality compared with the state-of-the-art methods.

Citation: Yang Y, Liu Q, Ji R, Gao Y (2012) Dynamic 3D Scene Depth Reconstruction via Optical Flow Field Rectification. PLoS ONE 7(11): e47041. doi:10.1371/journal.pone.0047041

Editor: Teresa Serrano-Gotarredona, National Microelectronics Center, Spain

Received: June 13, 2012; **Accepted:** September 7, 2012; **Published:** November 9, 2012

Copyright: © 2012 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by National Natural Science Foundation of China (NSFC) (No.61170194 and 61202301) and International S&T Cooperation Program of China (ISTCP) (No.2012DFG12010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rrji@ee.columbia.edu (RRJ); kevin.gaoy@gmail.com (YG)

† These authors contributed equally to this work.

Introduction

Depth maps are crucial for three-dimensional (3D) imaging [1] and displaying, and have been widely used in digital holography image processing [2,3], object reconstruction in integral imaging [4,5], 3D object retrieval [6–10] and tomographic phase microscopy [11]. Practically, high fidelity depth maps of a dynamic scene are calculated or captured in a temporal discrete manner due to the intensive computational complexity of depth reconstruction. For example, the widely used RGB-D [12,13] (e.g., Kinect) and ToF [14,15] camera can capture the depth map in video-rate but with low resolution (e.g., 320×240 pixels). The devices have challenges to capture depth maps for dynamic scene in video-rate with higher resolution (e.g., standard definition or even higher). In many 3D applications, it is noted that higher capture-rate depth map sequence with higher resolution is required to better represent a dynamic scene [13,15–18].

In order to solve the problem, depth propagation algorithms [19–21] have been investigated to compensate the capture-rate to video-rate of depth maps in recent years. In these algorithms, it is assumed that the variation for a given dynamic scene is identical for both the depth and the color information of one viewpoint. Specifically, objects keep static in consequent color frames will not arouse depth value variation for these objects, and the depth value for the region containing static object also keeps static in depth map. On the other hand, motions in consequent color frames correspond to depth value variation in the same region. Therefore, the status (i.e., static or motive) of object in consequent color

frames can be used to describe the depth value variation in depth map. Motion vector is widely applied to describe the motion status of objects, and it can be obtained with pixel-, block- or region-wise and with different accuracy. For the case of depth propagation, the pixel-wise motion vectors (PMVs) in consequent color frames with high accuracy can be mapped to depth maps also with high accuracy. Based on the assumption, low capture-rate depth maps can be compensated to video-rate. In this case, the captured and high resolution depth maps can be treated as key frames, and the depth information in to-be-reconstructed depth maps is propagated from the key frame by the obtained PMVs.

The main problem for depth propagation is that it is very challenging to obtain accurate PMVs for the occlusive or low textural regions, although PMV can have high accuracy in other regions. Inaccurate PMVs in these regions may lead to ambiguities and holes in the reconstructed depth maps, which decrease the reconstruction quality significantly. Variety of methods were proposed to improve the quality of reconstructed depth maps. For example, manual marking from users on potential problem regions can improve the quality of reconstructed depth maps significantly [20], but this method is not applicable in many automatic processing systems. Some filters have been applied in post-processing of reconstructed depth map, such as bilateral filter [19,21], discontinuity analysis and interpolation [22] and inpainting [23]. These filters usually result in inevitable and undesired blurs for depth maps, and these artifacts are unfavorable for 3D dynamic scene representation. In order to solve this

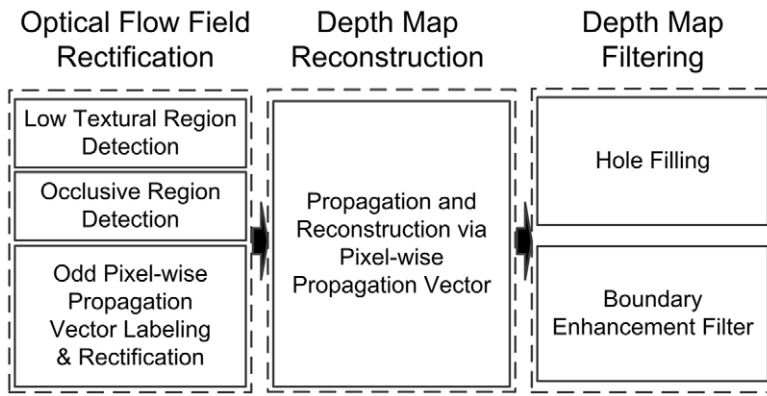


Figure 1. The schematic overview of our method.
doi:10.1371/journal.pone.0047041.g001

problem, we propose to rectify the optical flow field before propagation rather than rectify the depth results after propagation. The quality of the reconstructed depth is improved by PMVs rectification since global filtering has been avoided in this method. Furthermore, a boundary enhancement filter is proposed to refine the edges of the reconstructed depth maps. The main contributions of our work are three-fold: (1) propose a depth propagation scheme based on optical flow field rectification, in which high accurate PMVs can be obtained to improve the precision of propagation and the quality of reconstructed depth map, (2) propose an adaptive occlusive and low textural regions detection and rectification method for PMVs, and (3) propose a boundary enhancement filter to refine the reconstructed depth map.

Materials and Methods

Overview of the Proposed Rectification Method

In this work, we reconstruct depth maps for dynamic 3D scene in video-rate by propagation. The PMVs among consequent color images describe the temporal correlations in pixel-wise, and can be applied in propagation from the key depth map to the consequent vacant depth maps. Therefore, the quality of reconstructed vacant depth map highly depends on the precision of obtained PMV. Originally, PMVs can be calculated by traditional optical flow algorithms or motion estimation methods. However, the precision of obtained PMV decreases in several regions, for example, the

occlusive or low textural regions. In these regions, less information is available to the matching procedure in determining PMV, and thus errors in obtained PMV are inevitable. These errors will result in unreliable PMVs for depth reconstruction. In order to improve the quality of reconstructed depth maps, the obtained PMVs are rectified before propagation and reconstruction in our work.

Figure 1 shows the schematic overview of our proposed scheme. As we mentioned above, a rectification on PMVs is performed after PMVs have been obtained by optical flow algorithm. The rectification is performed to solve the problem caused by the occlusive and low textural regions. After that, depth information in key depth map is propagated to vacant depth maps through the rectified PMVs for depth map reconstruction. Finally, a depth map filtering will be performed finally to improve the quality of reconstructed depth. The details of each step of our proposed scheme will be given in following subsections.

Optical Flow Field Rectification

The unreliable PMVs usually occur in occlusive or low textural regions as aforementioned. Therefore, these regions should be detected properly at first.

The texture complication is an important clue for the goal of detection. Usually, the texture complication keeps consistency with the variation of pixel value, so that it can be represented by standard deviation of pixel values. The Heaviside step function is a unit step function, and it can be denoted by

$$y = \lim_{k \rightarrow \infty} \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1}(kx) \right) \quad (1)$$

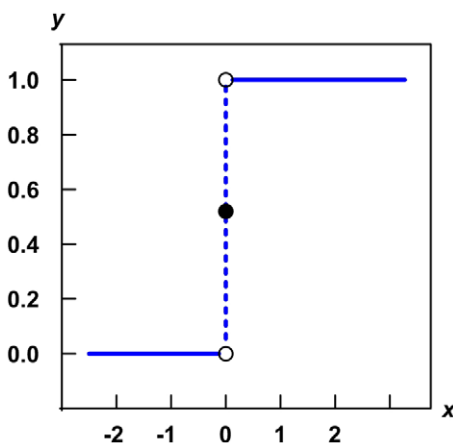


Figure 2. Heaviside step function.
doi:10.1371/journal.pone.0047041.g002

This function is always applied in the mathematics of control theory and signal processing. This function is a discontinuous function whose value is 0 for negative argument and 1 for positive argument. As shown by Figure 2, the function represents a signal that switches on at a specified time (usually triggered by a threshold) and stays switched on indefinitely. In order to differentiate the pixel $\mathbf{X}=(x,y)$ in color image, we propose a binary decision function $\zeta(\Omega_{\mathbf{X}})$ in form of the Heaviside step function to determine whether a region $\Omega_{\mathbf{X}}$ is the low textural region by

$$\zeta(\Omega_{\mathbf{X}}) = \lim_{k \rightarrow \infty} \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} k \left(\sigma \left(I_{\mathbf{v} \in \Omega_{\mathbf{X}}} - I_{\mathbf{X}} \right) - \varepsilon_{\Omega} \right) \right) \quad (2)$$

where $\Omega_{\mathbf{X}}$ is the neighboring pixel set centered at \mathbf{X} , \mathbf{Y} is a pixel in $\Omega_{\mathbf{X}}$, $I_{\mathbf{X}}$ is the gray value for \mathbf{X} , $\sigma(\cdot)$ is standard deviation operator for a set, and ε_{Ω} is a threshold for texture. According to the definition of Heaviside step function, the value of $\zeta(\Omega_{\mathbf{X}})$ corresponds to a binary decision for textural region detection. $\zeta(\Omega_{\mathbf{X}})=0$ indicates the pixel \mathbf{X} is surrounded by textures, but in low textural region when $\zeta(\Omega_{\mathbf{X}})=1$.

Similarly, we also propose a binary decision function $\rho(\mathbf{v})$ to determine whether a pixel \mathbf{X} is occluded

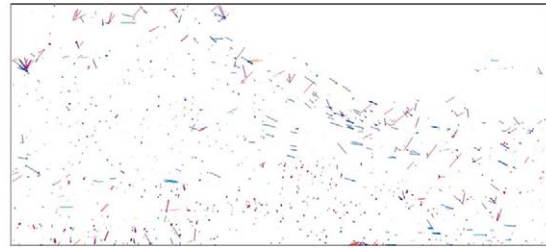
$$\rho(\mathbf{v}) = \lim_{k \rightarrow \infty} \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} k \left((I_{\mathbf{X}+\mathbf{v}} - I_{\mathbf{X}})^2 - \varepsilon_I \right) \right) \quad (3)$$

where \mathbf{v} is the PMV on \mathbf{X} , ε_I is a threshold for occlusions. $\rho(\mathbf{v})=1$ is for the occluded pixel \mathbf{X} , and $\rho(\mathbf{v})=0$ is for the visible one. In determining the occlusive or low textural regions, a smaller threshold is related to accurate decisions and stable performances on different test materials, while increase the computation complexity and unfavorable for implementations. On the other hand, larger threshold is benefit for implementations, but decrease the accuracy of decision and have unstable performances. We will discuss the parameter settings for thresholds ε_{Ω} and ε_I in the section of experiments in details.

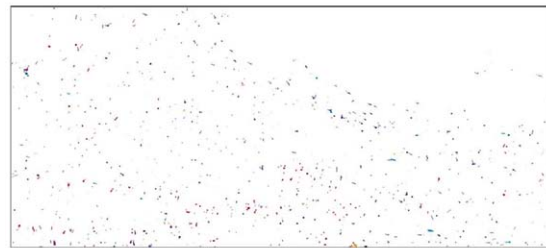
Based on $\zeta(\Omega_{\mathbf{X}})$ and $\rho(\mathbf{v})$, we can know the status of pixel \mathbf{X} and its surroundings, and make appropriate rectifications and operations on them. There are several cases for different combinations of decisions caused by the binary value of $\zeta(\Omega_{\mathbf{X}})$ and $\rho(\mathbf{v})$. For the first case, when the pixel \mathbf{X} is occluded by other objects (i.e., $\rho(\mathbf{v})=1$), the vector \mathbf{v} for \mathbf{X} is an error PMV since actually no corresponding pixel can be found for \mathbf{X} , no matter \mathbf{X} is surrounded by textures or not. In this case, it is not an easy way to predict a proper value for \mathbf{v} directly from neighboring vectors. Therefore, we mark \mathbf{X} with a label $label(\mathbf{v})$, treat \mathbf{v} as unreliable and process the depth value for \mathbf{X} after the depth map has been reconstructed. Then for the second case, the pixel \mathbf{X} is visible and surrounded by textures (i.e., $\rho(\mathbf{v})=0$ and $\zeta(\Omega_{\mathbf{X}})=0$). Texture information is benefit for accurate optical flow calculation, and thus the vector \mathbf{v} can be treated as reliable and accurate. Finally, the pixel \mathbf{X} is visible but in a low textural region (i.e., $\rho(\mathbf{v})=0$ and $\zeta(\Omega_{\mathbf{X}})=1$). Low textural region can cause pixel-wise ambiguous vectors in optical flow calculation. These unreliable PMVs of ambiguous are usually odd when comparing with neighboring vectors, as can be found in Figure 3(a). In this case, the unreliable odd vectors can be processed by an average filtering with the neighboring vectors. We summarize the above processing as a condition function



(a)



(b)



(c)

Figure 3. Result of rectification on optical flow field. (a) The almost static scene. (b) The obtained optical flow field without rectification. (c) The rectified optical flow field. doi:10.1371/journal.pone.0047041.g003

$$\Psi(\mathbf{v}) = \begin{cases} label(\mathbf{v}) & \rho(\mathbf{v}) = 1 \\ \mathbf{v} & \rho(\mathbf{v}) = 0 \& \zeta(\Omega_{\mathbf{X}}) = 0 \\ avg \left(\begin{matrix} \mathbf{v} \\ \mathbf{v} \in \Omega_{\mathbf{X}} \end{matrix} \right) & \rho(\mathbf{v}) = 0 \& \zeta(\Omega_{\mathbf{X}}) = 1 \end{cases} \quad (4)$$

where $label(\mathbf{v})$ is a mark on \mathbf{X} that \mathbf{v} is reserved for the next step processing, $avg(\cdot)$ is an average operator on a set.

Based on the above occlusive and low textural region detection and rectification, a part of the obtained unreliable (i.e., odd) PMVs

Table 1. Challenges in dynamic 3D scene materials.

Sequence Name	Challenge
Dancer	Full tiny but orderless textures: miss-matches in motion estimation.
Balloon	Illumination variation and focus-light rotation: non-zero MV for static region.
Lovebird1	Zooming out movement: overlap of neighboring MVs.

doi:10.1371/journal.pone.0047041.t001

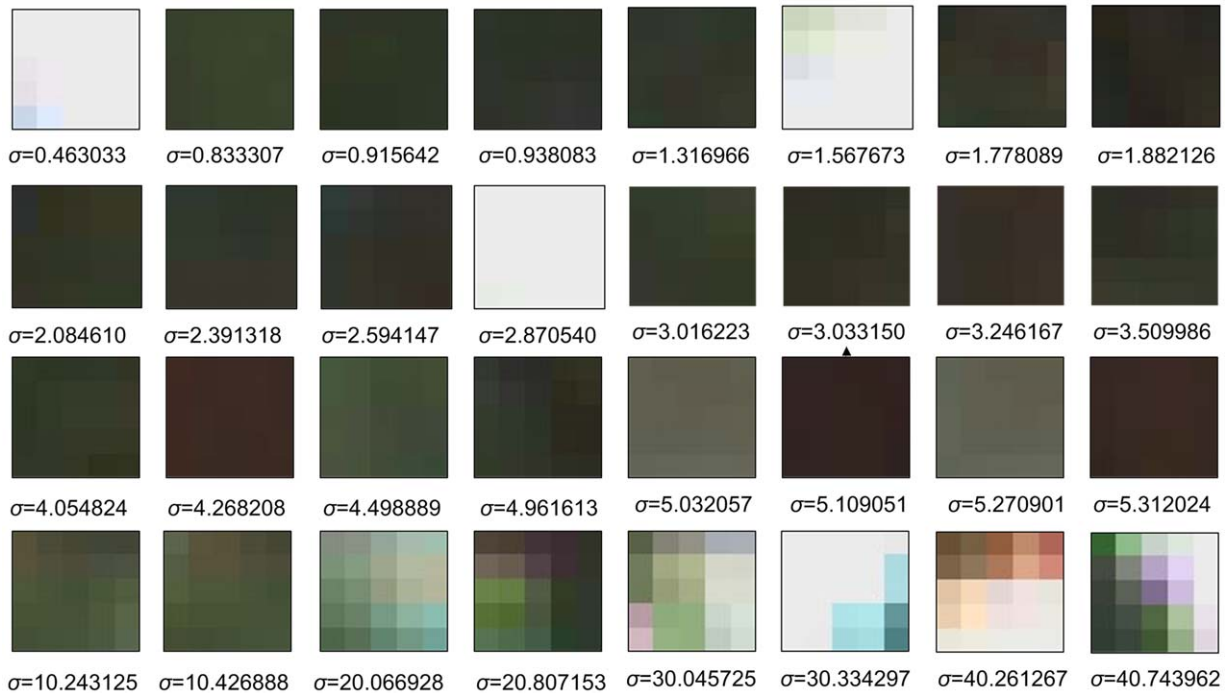


Figure 4. Texture analysis for color image for different $\sigma(I_{v \in \Omega_x} - I_x)$.
doi:10.1371/journal.pone.0047041.g004

can be rectified effectively, and the error PMVs from occlusive regions are reserved for later processing. Therefore, the accuracy of PMVs is improved.

Depth Map Reconstruction

Figure 3 provides the results of occlusive and low textural region detection and rectification. The results are obtained from two consequent color images in “Lovebird1” from MPEG [24]. The optical flow field is obtained between $I_t \rightarrow I_{t+1}$ (i.e., from the previous image to the current one), and errors will occur in low textural regions in background and occlusions around object boundaries. Figure 3(a) shows the static scene with occlusive and low textural regions, and Figure 3(b) is the obtained optical flow field based on the given static scene of Figure 3(a) where many unreliable PMVs can be found. Figure 3(c) shows the result that Figure 3(b) processed by the operator in Equation 4. It can be found that most of the odd PMVs have been rectified.

After that, the vacant depth map $(D^{t+1})_{n \times m}$ at time $t+1$ can be propagated and reconstructed via the obtained optical flow field $\Psi(\mathbf{v})_{n \times m}$ from the previous depth map as

$$(D_{\mathbf{X}+\Psi(\mathbf{v})}^t)_{n \times m} \rightarrow (D_{\mathbf{X}}^{t+1})_{n \times m} \quad (5)$$

A depth map sequence that synchronized with the color frames can be reconstructed by Equation 5. This depth map sequence is with high resolution in video-rate. The depth information in vacant time slot is propagated from the key depth map, where the depth information is reliable. However, the processing on the reconstructed depth maps is not finish yet. As denoted by Equation 4, the PMVs for occlusive region is labeled and reserved for post-processing, and the regions that reserved will be a hole without depth information in the reconstructed depth map. On the other hand, ambiguous PMVs is inevitable in optical flow algorithms.

These PMVs also can result in holes. Therefore, a depth map filtering for post-processing is necessary to improve the quality of reconstructed depth maps.

Depth Map Filtering

As mentioned above, the reconstructed depth maps may contain holes due to the labeling operation in Equation 4, and ambiguous PMVs. For the ambiguous PMVs on pixel \mathbf{X} , the missing depth value $d_{\mathbf{X}}$ is very close to its spatial neighbors $\Omega_{\mathbf{X}} \subset (D^{t+1})_{n \times m}$, so that a median filter is applied. The operation can be denoted by

$$d_{\mathbf{X}} = \text{median}\{d_{\mathbf{Y}} | \mathbf{Y} \in \Omega_{\mathbf{X}}\} \quad (6)$$

For the hole caused occlusions and marked in Equation 4, the pixel \mathbf{X} is marked by $label(\mathbf{v})$. In this case, the missing depth value $d_{\mathbf{X}} = d_{label}$ on pixel \mathbf{X} can be joint predicted by the depth value around the hole and the region where the PMV pointing to. We propose a depth value predictor as

$$d_{label} = \alpha \delta(\Omega_{label}) + (1 - \alpha) \delta(\Omega_{label+\mathbf{v}}) \quad (7)$$

where $\alpha \in [0, 1]$ is normalized by the norm of \mathbf{v} , $\delta(\cdot)$ is an in-painting operator [25].

However, in-painting on depth map will bring noticeable blur effect, especially when the hole crosses the boundaries of high contrast edges. Usually, depth map with blurred boundaries results in a failure on foreground-background separation [26]. Therefore, an object boundary enhancement filter (BEF) is further proposed

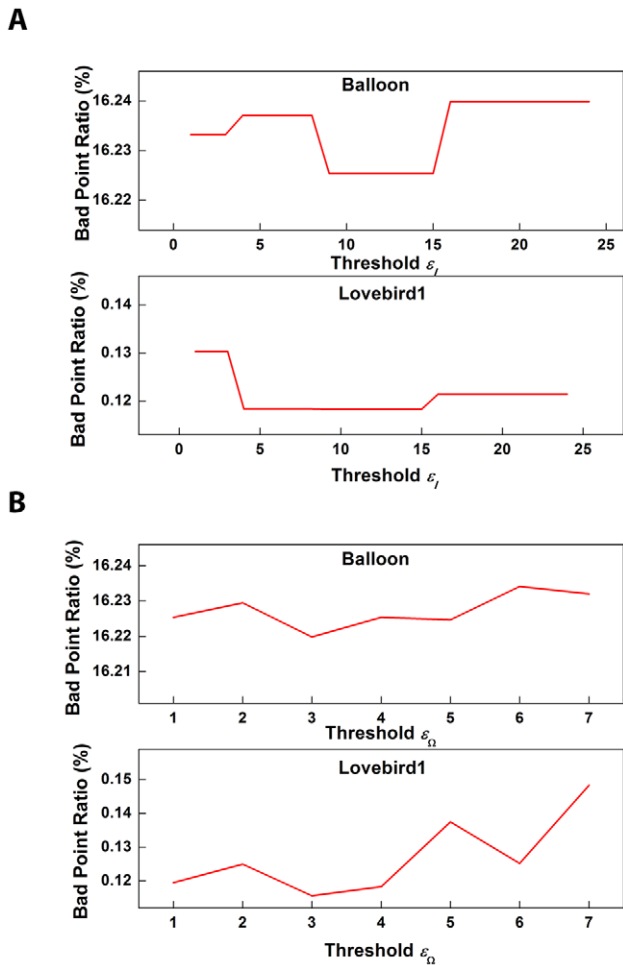


Figure 5. Parameter analysis for ϵ_Ω and ϵ_I . (a) The Bad Point Ratio with different ϵ_I selection when ϵ_Ω is fixed as 4. (b) The Bad Point Ratio with different ϵ_Ω selection when ϵ_I is fixed as 9. doi:10.1371/journal.pone.0047041.g005

$$h(D(\mathbf{X})) = k_d^{-1}(D(\mathbf{X})) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{f}(\xi)c(\xi, D(\mathbf{X}))d\xi \quad (8)$$

where

$$D(\mathbf{X}) = \arg \max_{Y \in \Omega_X} \text{freq}(D(\mathbf{Y})) \quad (9)$$

is the depth value that most frequent appearing in Ω_X , $h(\cdot)$ is a bilateral filter defined in [27]. $\text{freq}(\cdot)$ is a statistical function that count the appearing frequency of each element in a data set. For example, suppose we have a data set $A = \{a, a, b\}$, the result of $\text{freq}(A)$ will be $\{\{a, 2\}, \{b, 1\}\}$. Therefore, we can further have $\arg_A \max \text{freq}(A) = a$.

Equations 8 and 9 can smooth the depth map, and the object boundary can be sharper.

Dynamic 3D Scenes Materials

Dynamic 3D scenes contain a video-rate color image sequence that record the motion, color and texture information of this scene. Furthermore, a high resolution depth map sequence is also

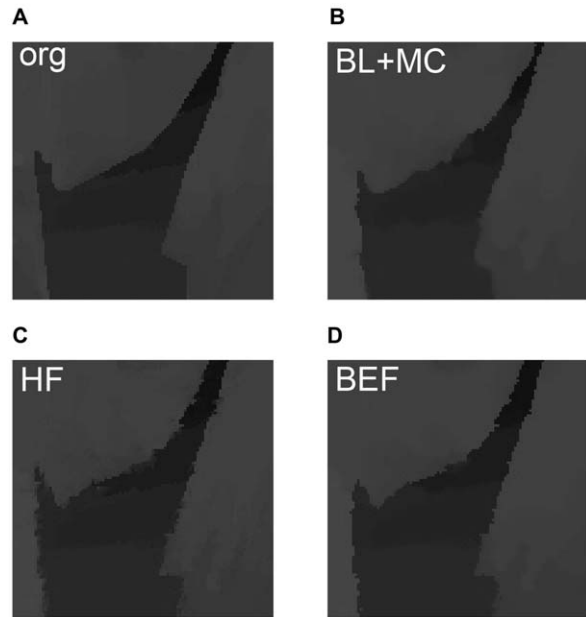


Figure 6. Subjective results on depth images for different filters. (a) Original depth map to be reconstructed. (b) Result of method in [19]. (c) Obtained depth map with hole-filling but without boundary enhancement filter. (d) Result of the proposed method. doi:10.1371/journal.pone.0047041.g006

captured to record the 3D space information for all visible objects. As we have mentioned above, high resolution depth map cannot be captured by RGB-D or ToF cameras in video-rate that synchronically with the color image sequence so far. Recently, MPEG released their standard test sequences for dynamic 3D scenes with high resolution (more than standard definition) and high frame-rate, including color images and depth maps [24]. The color images were captured by cameras, but the depth maps were not captured but calculated by stereo matching and even manual labeling. The quality of depth map obtained through this way was assumed with the best quality to be obtained.

The dynamic 3D scene materials named as Undo Dancer, Lovebird1, and Balloon will be used to testify our proposed algorithm. The captured color images and calculated depth maps

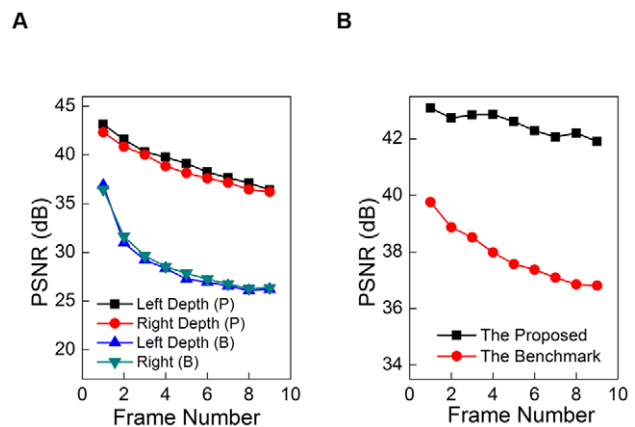


Figure 7. Objective results of our method. (a) Reconstructed quality for 9 consequent depth maps of left and right views for Lovebird1. (b) Quality for 9 synthesized correspondent virtual view images. doi:10.1371/journal.pone.0047041.g007

Table 2. Comparisons on Quality of Reconstructed Depth Maps and Synthesized Virtual Views (dB).

Test Sequence	Scheme	Left View	Right View	Virtual View	Gain in Left View	Gain in Right View	Gain in Virtual View
Dancer	Proposed	31.7396	31.7792	37.8609	5.7873	5.7809	0.7355
	[19]	25.9523	25.9983	37.1254			
Lovebird1	Proposed	39.2648	38.6206	41.8520	10.5336	9.6364	3.9791
	[19]	28.7312	28.9842	37.8729			
Balloon	Proposed	32.4155	31.0498	42.5151	8.6305	7.9726	4.5184
	[19]	23.7850	23.0772	37.9967			

doi:10.1371/journal.pone.0047041.t002

of these materials are selected from [24]. These materials are with different challenges in depth reconstruction, as listed in Table 1.

Results and Discussion

Experiment Arrangements

Experiments are arranged in four parts, including a discussion on thresholds in Equations 2 and 3, subjective and objective quality comparisons on depth reconstruction between our algorithm and the benchmark state-of-the-art method in [19], and finally an objective quality comparison on the dynamic 3D scene representation. In [19], the PMVs between consequent color images are not processed before propagation. Instead, a bilateral filter was applied for propagation, and errors in reconstructed depth were processed by motion compensation.

In our experiment, the depth map at $t=0$ is selected from the given depth map sequence and treated as key depth map. The consequent depth maps in material are treated as vacant, and they will be propagated and reconstructed by our proposed algorithm and the benchmark method with the help of the key depth map. The consequent depth maps in material will be used as anchor for the reconstructed depth in objective quality evaluation.

On the Parameters ε_{Ω} and ε_I

From Equations 2 and 3 we can see that there are two thresholds ε_{Ω} and ε_I in our formulation. They modulate the number of the pixels of occluded or low textural, and thus the final quality of output propagated depth maps. These parameters (i.e., thresholds) are usually used in pixel classification. For parameter ε_{Ω} , it is a real number varies in $[0, \infty)$, and it determines the number of low textural pixels. If ε_{Ω} tends to be infinite, all pixels in image will be determined as low textural ones no matter how many textures around them. According to our proposed scheme, spatial filter (i.e., average filtering in Equation 4) is applied on the low textural pixels. Depth information for textural pixels can then be erased by this filter, and thus the accuracy of obtained depth map will be degraded. For parameter ε_I , it is also a real number varies in $[0, \infty)$, and it determines the number of occluded pixels. Furthermore, Equation 3 is performed on two corresponding pixels that related by vector \mathbf{v} . The accuracy of \mathbf{v} can be represented by the deference of $I_{\mathbf{X}+\mathbf{v}} - I_{\mathbf{X}}$ and checked by ε_I .

Figure 4 demonstrates a texture analysis of $\Omega_{\mathbf{X}}$ on one color image in test material Lovebird1 when $\sigma(I_{\mathbf{V}\in\Omega_{\mathbf{X}}} - I_{\mathbf{X}})$ is changing. Texture in image can be treated as wave variation in signal. According to the definition of information entropy, more information is contained in $\Omega_{\mathbf{X}}$ when the signal varying sharply. When considering the matching operation in optical flow calculation, more information in $\Omega_{\mathbf{X}}$ is helpful to obtain higher

accurate and reliable \mathbf{v} . Therefore, the parameter of ε_{Ω} is also a threshold to distinguish reliable and unreliable \mathbf{v} . Figure 4 shows that the region $\Omega_{\mathbf{X}}$ can be clearly classified to low textural region when $\sigma(I_{\mathbf{V}\in\Omega_{\mathbf{X}}} - I_{\mathbf{X}}) < 6$, or otherwise, apparent textures are visible in $\Omega_{\mathbf{X}}$.

Based on the texture analysis in Figure 4, the parameter settings for ε_{Ω} and ε_I can be solved. Figure 5 demonstrate the performance curves (i.e., Bad Point Ratio) with respect to the variation of ε_{Ω} and ε_I . In Figure 5 (a), we fix ε_{Ω} to be 4 and vary ε_I from 1 to 20. We can see that Bad Point Ratio drops to minimum point when ε_I is 9. After that, in Figure 5 (b), we fix ε_I to be 9 and vary ε_{Ω} from 1 to 7. It can be found that Bad Point Ratio varies slightly for the parameter ε_{Ω} , but the curve is increasing when ε_{Ω} becomes larger. Therefore, we select ε_{Ω} to be 3 to obtain a relative smaller Bad Point Ratio, indicating higher accuracy of depth map.

Subjective Results for Depth Reconstruction

Figure 6 gives comparison results of subjective experiments. Each subfigure provides an enlarged part, and details the difference between our algorithm and the method in [19]. Figure 6(a) is the original depth map that selected in materials, and it serves as benchmark and is treated as absent in depth reconstruction. Figure 6(b) marked by “BL+MC” is obtained by method in [19], and it shows definite geometric distortion around the regions of moving object boundary. The phenomenon is a result of temporal bilateral-filtering. On the contrary, our algorithm detects the occlusive and low textural region, and processes these regions according to their types before depth propagation and reconstruction. Figure 6(c) marked by “HF” is the reconstructed depth map by using the optical flow field $\Psi(\mathbf{v})_{n \times m}$ in Figure 3(c) with Equations 5, 6 and 7. As we mentioned above, the blurring effect is occurred around object boundaries. Figure 6(d) marked by “BEF” is the result obtained by our proposed method. The operation difference between Figures 3(c) and (d) is the BEF, which is processed by Equation 8 and 9. It can be found that the blurring effect is removed and the boundary around object is sharper.

Objective Results for Depth Reconstruction

The objective quality comparison is measured by the peak signal-noise ratio (PSNR) from the reconstructed and corresponding existing depth maps from test materials. In the comparisons, higher PSNR indicates higher accuracy and better performance. Figure 7 and Table 2 provide the quantitative results. We can see that high precision of depth propagation is benefit in high quality of depth reconstruction, and the quality of reconstructed depth map of our method (labeled with “P”) is more than 8 dB better than the benchmark (labeled with “B”). However, errors (e.g.

distortions around boundary) will also be propagated as shown in Figure 6(b). Therefore, the quality of reconstructed depth map will drop down along with longer distance propagation. As for the results given in Figure 7(a), we reconstruct 9 consequent depth maps for Lovebird1 for both left and right views respectively. Figure 7(a) shows that the quality of the 1st depth map reconstructed by the benchmark is comparable with that of the 9th by our method, indicates the higher quality of our method. On the other hand, Table 2 lists the average quality results on 9 reconstructed depth maps for three test sequences. It is obvious that our method has at least 5 dB gains on depth reconstruction, which is due to the rectification on optical flow field. On the other hand, BEF is used to eliminate the blur effect around boundary, and it will also benefit the quality.

Results in Dynamic 3D Scene Representation

Dynamic 3D scene representation is measured by the objective quality of virtual view synthesis. Virtual view is an important application in 3D computer vision when color images and the corresponding depth maps are both available for a dynamic 3D scene [13,17]. Better quality of depth maps can yield high quality

of virtual view, and have better performance in dynamic 3D scene representation.

We use the reconstructed depth map for synthesis by VSRS software [28], which is a common test platform. The results are also given in Figure 7(b) and Table 2. Our method achieves 0.7 to 4.5 dB gains on PSNR for all the test materials. On the other hand, the accuracy of reconstructed depth map from the benchmark will be greatly affected by filter-based propagation. The distortion results in synthesis distortions.

Summary of Results

In sum of the above quantitative comparisons, the proposed algorithm can achieve more accurate depth reconstruction on all test sequences with different challenges, including global motion and local motion, or dynamic scene that captured in natural environment and generated by computer graphics.

Author Contributions

Conceived and designed the experiments: RRJ YG. Performed the experiments: QL YY. Analyzed the data: YY RRJ. Wrote the paper: YY QL YG.

References

- Gao Y, Dai Q, Zhang N (2010) 3D model comparison using spatial structure circular descriptor. *Pattern Recognition* 43: 1142–1151.
- Smith ZJ, Chu I K, Espenson AR, Rahimzadeh M, Gryshuk A, et al. (2011) Cell-phone-based platform for biomedical device development and education applications. *PLoS ONE* 6: e17150. doi:10.1371/journal.pone.0017150.
- Pavillon N, Kuhn J, Moratal C, Jourdain P, Depeursinge C, et al. (2012) Early cell death detection with digital holographic microscopy. *PLoS ONE* 7: e30912. doi:10.1371/journal.pone.0030912.
- Jung JH, Hong K, Park G, Chung I, Park JH, et al. (2010) Reconstruction of three-dimensional occluded object using optical flow and triangular mesh reconstruction in integral imaging. *Optics Express* 18: 26373–26387.
- Lee JJ, Lee BG, Yoo H (2011) Depth extraction of three-dimensional objects using block matching for slice images in synthetic aperture integral imaging. *Applied Optics* 50: 5624–5629.
- Gao Y, Wang M, Zha Z, Tian Q, Dai Q, et al. (2011) Less is more: Efficient 3D object retrieval with query view selection. *IEEE Transactions on Multimedia* 11: 1007–1018.
- Wang M, Ni B, Hua XS, Chua TS (2012) Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Computer Survey* 44: 1–24.
- Gao Y, Wang M, Tao D, Ji R, Dai Q (2012) 3D object retrieval and recognition with hypergraph analysis. *IEEE Transactions on Image Processing* 21: 4290–4303.
- Wang M, Hong R, Yuan XT, Yan S, Chua TS (2012) Movie2comics: Towards a lively video content presentation. *IEEE Transactions on Multimedia* 14: 858–870.
- Gao Y, Tang J, Hong R, Yan S, Dai Q, et al. (2012) Camera constraint-free view-based 3D object retrieval. *IEEE Transactions on Image Processing* 21: 2269–2281.
- Choi W, Fang-Yen C, Badizadegan K, Dasari RR, Feld MS (2008) Extended depth of focus in tomographic phase microscopy using a propagation algorithm. *Optics Letters* 33: 171–173.
- Leyvand T, Meekhof C, Wei YC, Sun J, Guo B (2011) Kinect identity: Technology and experience. *Computer* 44: 94–96.
- Henry P, Krainin M, Herbst E, Ren X, Fox D (2010) Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In: *International Symposium on Experimental Robotics*. 1–6.
- Kolb A, Barth E, Koch R (2008) Tof-sensors: New dimensions for realism and interactivity. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 1–6.
- Cho JH, Kim SY, Ho YS, Lee K (2008) Dynamic 3d human actor generation method using a time-of-flight depth camera. *IEEE Transactions on Consumer Electronics* 54: 1514–1521.
- Liu Q, Yang Y, Ji R, Gao Y, Yu L (2012) Cross-view down/up-sampling method for multiview depth video coding. *IEEE Signal Processing Letters* 19: 295–298.
- Smolic A, Miller K, Merkle P, Fehn C, Kauff P, et al. (2006) 3d video and free viewpoint video - technologies, applications and mpeg standards. In: *International Conference on Multimedia and Exposition*.
- Fehn C (2004) Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv. In: *Stereoscopic Displays and Virtual Reality Systems XI, SPIE*. 93–104.
- Varekamp C, Barenbrug B (2007) Improved depth propagation for 2d-to-3d video conversion using key-frames. In: *European Conference on Visual Media Production*. 1–7.
- Yan X, Yang Y, Er G, Dai Q (2011) Depth map generation for 2d-to-3d conversion by limited user inputs and depth propagation. In: *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*. 1–4.
- Cao X, Li Z, Dai Q (2011) Semi-automatic 2d-to-3d conversion using disparity propagation. *IEEE Transactions on Broadcasting* 57: 491–499.
- Lee SB, Ho YS (2009) Discontinuity-adaptive depth map filtering for 3d view generation. In: *Proceedings of the 2nd International Conference on Immersive Telecommunications. IMMERSCOM '09*, 1–6.
- Ince S, Konrad J (2008) Occlusion-aware optical flow estimation. *IEEE Transactions on Image Processing* 17: 1443–1452.
- (2011) Call for proposals on 3d video coding technology. *ISO/IEC JTC1/SC29/WG11, Doc W12036*.
- Bertalmio M, Sapiro G, Caselles V, Ballester C (2000) Image inpainting. In: *SIGGRAPH*. 417–424.
- Muller K, Smolic A, Dix K, Merkle P, Kauff P, et al. (2008) View synthesis for advanced 3d video systems. *EURASIP Journal on Image and Video Processing* Volume 2008, Article ID 438148. doi:10.1155/2008/438148.
- Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: *International Conference Computer Vision*. 839–846.
- (2008) View synthesis tools for 3d video. *ISO/IEC JTC1/SC29/WG11, Doc M15851*.