

Research Article

Inaccuracy-Tolerant Sparse-to-Dense Depth Propagation for Semiautomatic 2D-to-3D Conversion

Hongxing Yuan 

School of Electronics and Information Engineering, Ningbo University of Technology, Ningbo 315211, China

Correspondence should be addressed to Hongxing Yuan; yuanhx@mail.ustc.edu.cn

Received 3 November 2017; Revised 3 June 2018; Accepted 24 June 2018; Published 8 August 2018

Academic Editor: Jit S. Mandeep

Copyright © 2018 Hongxing Yuan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Current semiautomatic 2D-to-3D methods assume that user input is perfectly accurate. However, it is difficult to get 100% accurate user scribbles and even small errors in the input will degrade the conversion quality. This paper addresses the issue with scribble confidence that considers color differences between labeled pixels and their neighbors. First, it counts the number of neighbors which have similar and different color values for each labeled pixels, respectively. The ratio between these two numbers at each labeled pixel is regarded as its scribble confidence. Second, the sparse-to-dense depth conversion is formulated as a confident optimization problem by introducing a confident weighting data cost term and the local and k -nearest depth consistent regularization terms. Finally, the dense depth-map is obtained by solving sparse linear equations. The proposed approach is compared with existing methods on several representative images. The experimental results demonstrate that the proposed method can tolerate some errors from user input and can reduce depth-map artifacts caused by inaccurate user input.

1. Introduction

3D videos have attracted more and more attention, providing an immersive realism visual experience by exploiting depth information [1]. With rapid advances in 3D display technologies, 3D content shortage has become one of the bottlenecks which restrict the development of entire 3D industry [2]. To remedy this issue, many 2D-to-3D conversion methods have been developed to convert existing 2D images/videos into 3D format by creating depth-maps [3]. Semiautomatic 2D-to-3D conversion can produce high-quality depth-maps from sparse user scribbles by using sparse-to-dense depth propagation. However, current methods assume that user scribbles are perfectly accurate [4–6], and depth quality degrades significantly when inaccurate scribbles are present. Figure 1 shows an experimental result where scribbles are partly inaccurate. As shown in Figures 1(d) and 1(e), existing methods generate visual artifacts around inaccurate labeled regions. To handle the inaccurate input, a confident sparse-to-dense propagation algorithm is introduced in this paper that

obtains accurate depth-maps even from erroneous user scribbles, as in Figure 1(f).

The proposed method is based on the observation that inaccurate input often occurs at or near object boundaries, and the number of correct scribbles is much larger than the number of incorrect ones. The rest of this paper is organized as follows. In Section 2, the related works about sparse-to-dense depth propagation for 2D-to-3D conversion are reviewed. The proposed method is described in Section 3. Experimental results are provided in Section 4. Finally, conclusion and future work are given in Section 5.

2. Related Work

2D-to-3D conversion algorithms can be categorized into manual methods, automatic methods, and semiautomatic methods. Manual methods can offer the highest quality conversion results but need per-pixel depth assignment precisely which is most time consuming and costly. Automatic methods infer depth information in image/video by

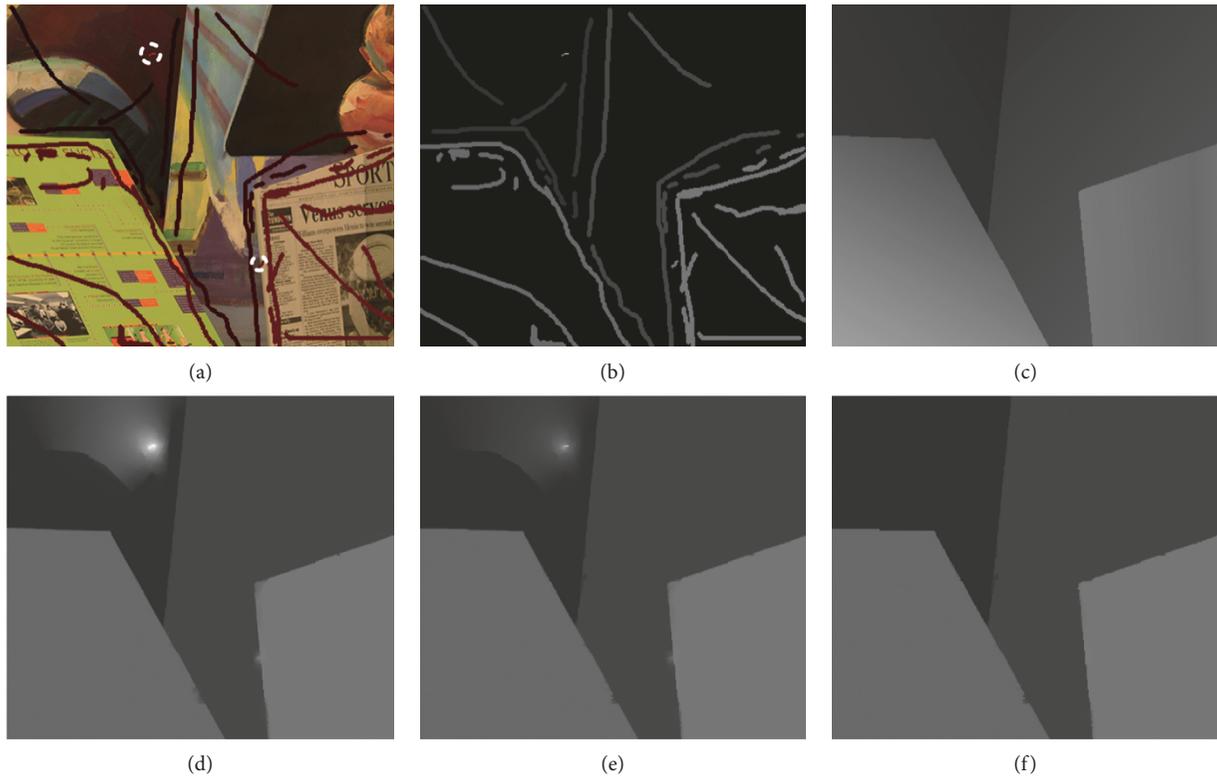


FIGURE 1: A sparse-to-dense depth propagation result with inaccurate input. (a) User scribbles (inaccurate scribbles are marked by the white circles). (b) Sparse depth-map extracted from user scribbles. (c) Groundtruth depth-map. (d) Depth-map generated by random walks [4]. (e) Depth-map generated by optimization method [6]. (f) Depth-map generated by the proposed method.

exploiting different depth perception cues such as motion, occlusion, vanishing points, defocus, and so on. Recently, with the popularity of deep learning, many neural networks have been proposed for automatic depth estimation [7–9]. However, existing automatic methods can generally provide a limited 3D effect due to ambiguities between depth and perception cues [2]. Semiautomatic methods are the most widely used schemes for 3D content creation, since they can balance conversion quality and production cost. The core step of semiautomatic methods is sparse-to-dense depth propagation on key frames, in which dense depth-maps are estimated from user-assigned sparse depth values. The conversion quality largely depends on the accuracy of depth-maps at key frames. Thus, this paper mainly focuses on sparse-to-dense depth propagation for semiautomatic 2D-to-3D conversion.

Phan and Androustos [10] combine random walks (RW) with graph cuts (GC) for sparse-to-dense depth estimation, but incorrectly segmented object boundaries provided by GC may degrade depth quality. Rzeszutek and Androustos [11] use the domain transform filter to propagate sparse labels throughout an image, but it may smooth out depth edges. Iizuka et al. [12] utilize superpixel-based geodesic distance weighting interpolation and optimization-based edge-preserving smooth to compute dense depth from user scribbles. Similarly, Wu et al. [13] apply superpixel-based optimization method to obtain dense depth-maps from sparse

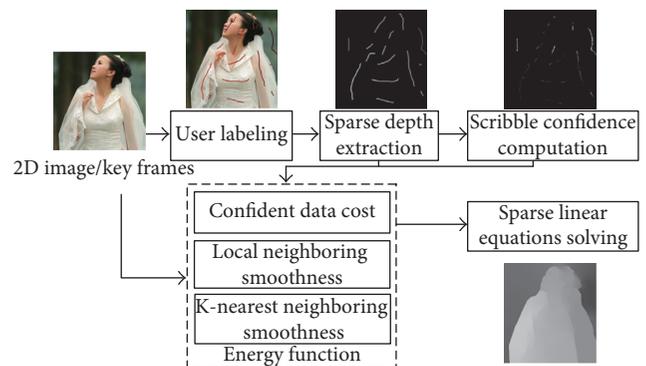


FIGURE 2: The flowchart of the proposed method.

input. However, these superpixel-based methods are affected by the performance of superpixel segmentation. Yuan et al. [14] propose a nonlocal RW algorithm to produce dense depth-map from user scribbles on single 2D image. Liang and Shen [15] further extend this scheme with the ability to process videos. However, RW-based methods cannot modify user-assigned labels, and erroneous input will degrade depth accuracy seriously. Lopez et al. [16] incorporate perspective and equality/inequality constraints into an optimization framework for dense depth estimation, but may add additional burden to user operations. Vosters and Haan [17]

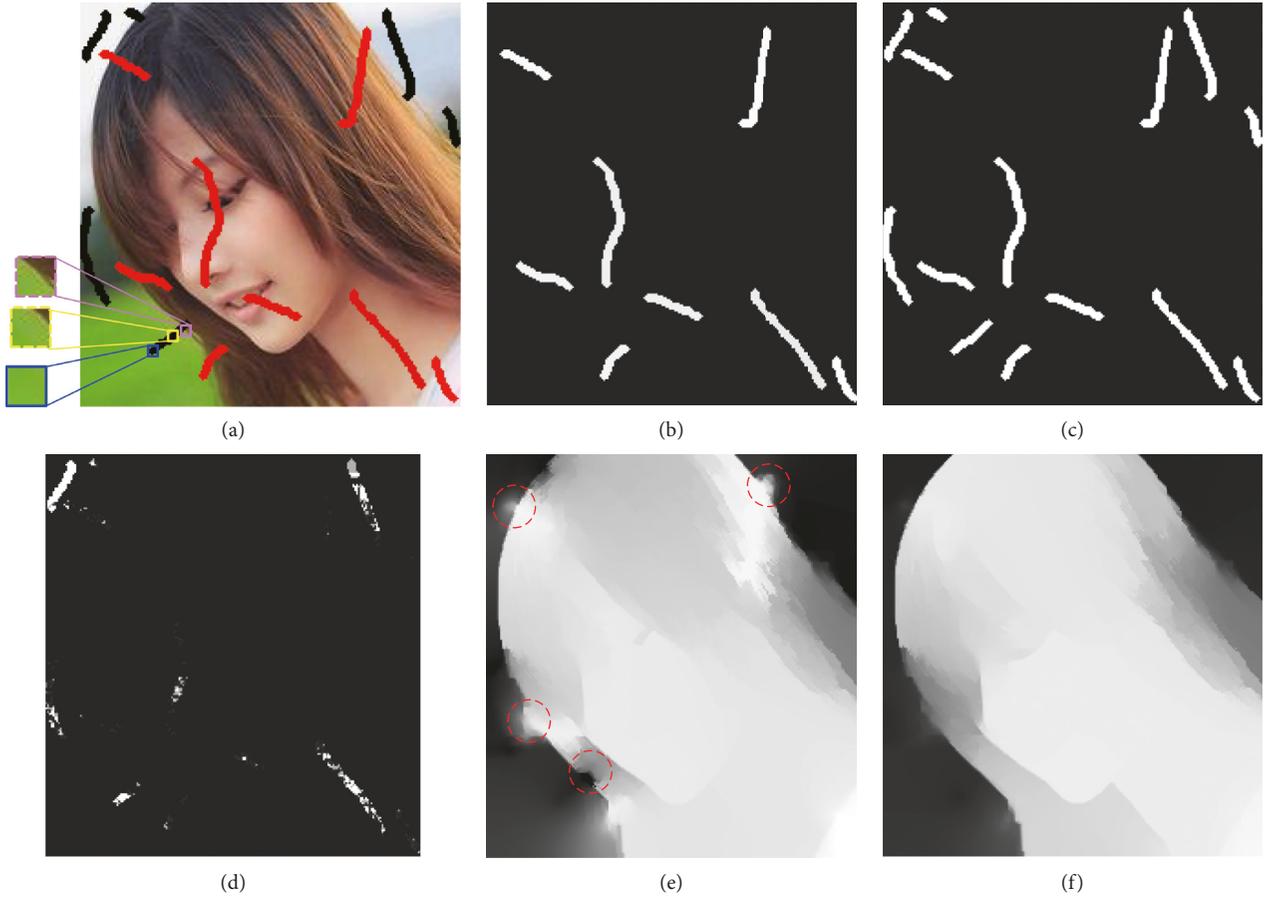


FIGURE 3: The demonstration of scribble confidence provided by the proposed method. (a) User scribbles. (b) Sparse depth. (c) Label mask. (d) Scribble confidence (the brighter the intensity, the higher the confidence of labeled pixels). (e) Depth-map of optimization method [6]. (f) Depth-map of confident optimization method.

propose a line scanning-based sparse-to-dense propagation method with low computation cost, but accuracy may be lost. Revaud et al. [18] use an edge-aware geodesic distance for sparse-to-dense optical flow interpolation, but the result is vulnerable to inaccurate input.

All of the above methods, however, do not account for the possibility of inaccurate scribbles. Thus, these methods can only give reliable results for accurate input. To address this issue, confidence of scribbles is calculated based on local color variation. There has been some recent works addressed on error-tolerant interactive image segmentation [19–21]. However, these methods are not well suited to 2D-to-3D conversion, since they mainly focus on foreground and background separation.

3. Method

As shown in Figure 2, the proposed method works as follows. Firstly, user draws sparse scribbles on 2D image/key frames, where brighter red marked regions are closer from the camera. Secondly, depth values at labeled pixels are extracted according to intensities of the scribbles. Thirdly, confidence of scribbles is calculated based on the color variation at labeled regions. Fourthly, an energy

function is built where scribble confidence is incorporated into the data cost. Finally, the energy function is minimized by solving a sparse linear system to obtain the dense depth-map.

3.1. Scribble Confidence. It can be found that pixels in accurate labeled regions often have similar color values, while erroneous input mainly appears at object boundaries with strong variations in color. Based on this observation, the scribble confidence is calculated using the following formula:

$$s_i = \begin{cases} \frac{\sum_{j \in \mathbf{N}(i)} \delta(\|\mathbf{I}_i - \mathbf{I}_j\|_2)}{\sum_{j \in \mathbf{N}(i)} \|\mathbf{I}_i - \mathbf{I}_j\|_0 + \varepsilon}, & \text{if } i \in \Omega, \\ 0, & \text{if } i \notin \Omega, \end{cases} \quad (1)$$

where \mathbf{I}_i denotes the Lab color values at pixel i . The reason for using Lab color space is that it takes into account human perception [22]. $\mathbf{N}(i)$ is the set of 8 neighbors of pixel i . $\delta(\cdot)$ is the Dirac delta function. $\|\cdot\|_0$ and $\|\cdot\|_2$ are L0 norm and L2 norm, respectively. ε is a small positive constant to prevent division by zero and set to 10^{-5} . Ω is the set of labeled pixels.

TABLE 1: SSIM comparison in the absence of inaccurate scribbles.

Image	Method					
	RW [4]	OPT [6]	HGR [10]	SOPT [13]	NRW [14]	Proposed
RGBZ_01	0.91	0.92	0.90	0.90	0.92	0.92
RGBZ_03	0.87	0.88	0.88	0.86	0.88	0.86
RGBZ_05	0.93	0.92	0.92	0.90	0.93	0.92
RGBZ_07	0.91	0.91	0.90	0.90	0.91	0.91
Average	0.91	0.91	0.90	0.89	0.91	0.91

The best SSIM at each row is shown in bold.

TABLE 2: PSNR comparison in the absence of inaccurate scribbles.

Image	Method					
	RW [4]	OPT [6]	HGR [10]	SOPT [13]	NRW [14]	Proposed
RGBZ_01	25.0	24.8	24.1	24.1	25.7	24.6
RGBZ_03	21.2	20.9	22.4	21.7	22.4	20.9
RGBZ_05	21.1	20.8	22.1	21.4	23.7	21.3
RGBZ_07	22.2	22.3	21.1	21.9	22.2	22.3
Average	22.4	22.2	22.4	22.3	23.5	22.3

The best PSNR (dB) at each row is shown in bold.

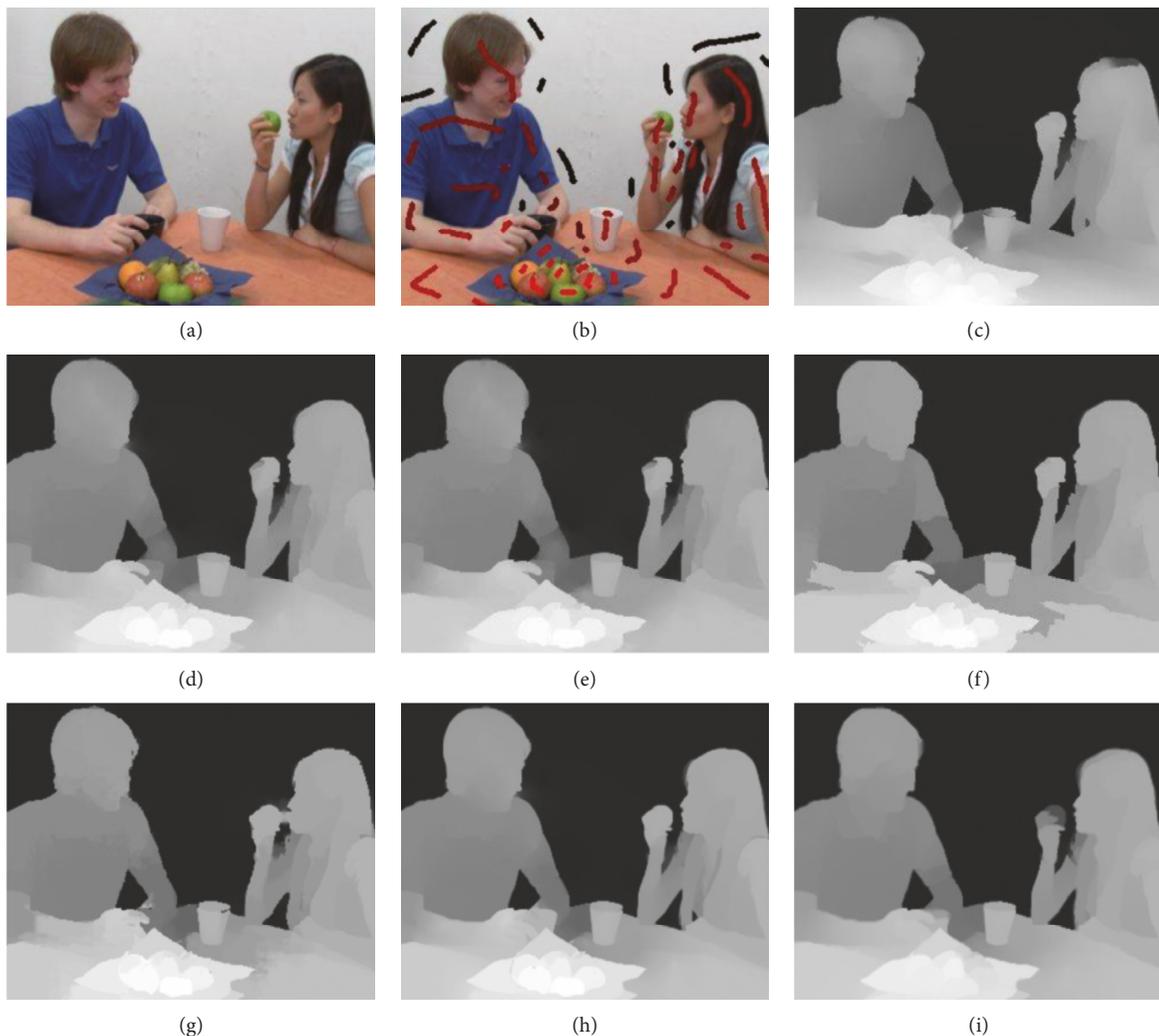


FIGURE 4: Results for RGBZ_01 in the absence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

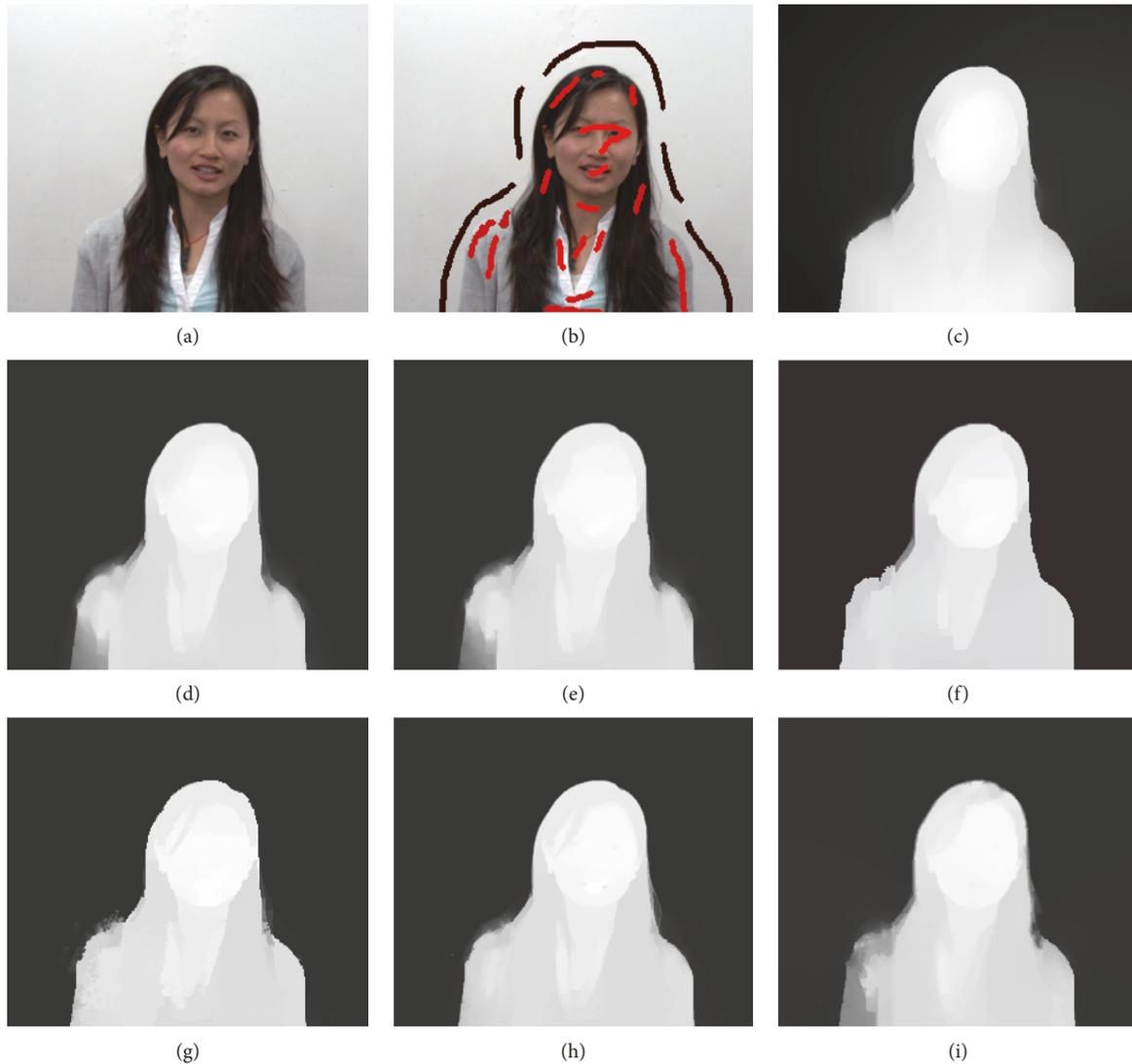


FIGURE 5: Results for RGBZ_03 in the absence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

It can be seen from formula (1) that the confidence of a labeled pixel is lower when its color difference between neighboring pixels is larger. Since inaccurate input is mainly located at or near object boundaries around which the color changes significantly, the proposed method can penalize inaccurate scribbles. One may question whether or not the confidence of correct scribbles is high. The confidence of labeled pixels at texture regions is indeed low, and correct labels at these regions will be mistaken as incorrect ones. However, the number of correct scribbles is much larger than the number of incorrect ones. Therefore, the impact on the accurate scribbles can be tolerated.

Figure 3 gives an example on how scribble confidence works. The confidence of inaccurate scribbles which move across the object boundaries is low, as shown in Figure 3(d). Current optimization method [6] generates visual artifacts around inaccurate labeled regions, as can be seen from regions within the red circles in Figure 3(e). These artifacts can be removed if scribble confidence is

incorporated into the optimization method, as shown in Figure 3(f).

As shown as the blue square in Figure 3(a), when user scribbles are inside objects, the color variation at labeled pixel between its neighbors is small, in which case the scribble confidence of the pixel at the center of the blue square is 0.6. When user scribbles approach object boundaries, the color variation at labeled pixel between its neighbors becomes larger, as shown as the yellow and pink squares in Figure 3(a). The scribble confidence of the pixel at the center of the yellow square is 0.3, while confidence of the center pixel of the pink square is 0.0. Since erroneous scribbles mainly appears at object boundaries, the proposed method can remove erroneous input by using color differences between labeled pixels and their neighbors.

3.2. Energy Function. Let n be the total number of pixels and w and h the image width and height in pixels, that is,

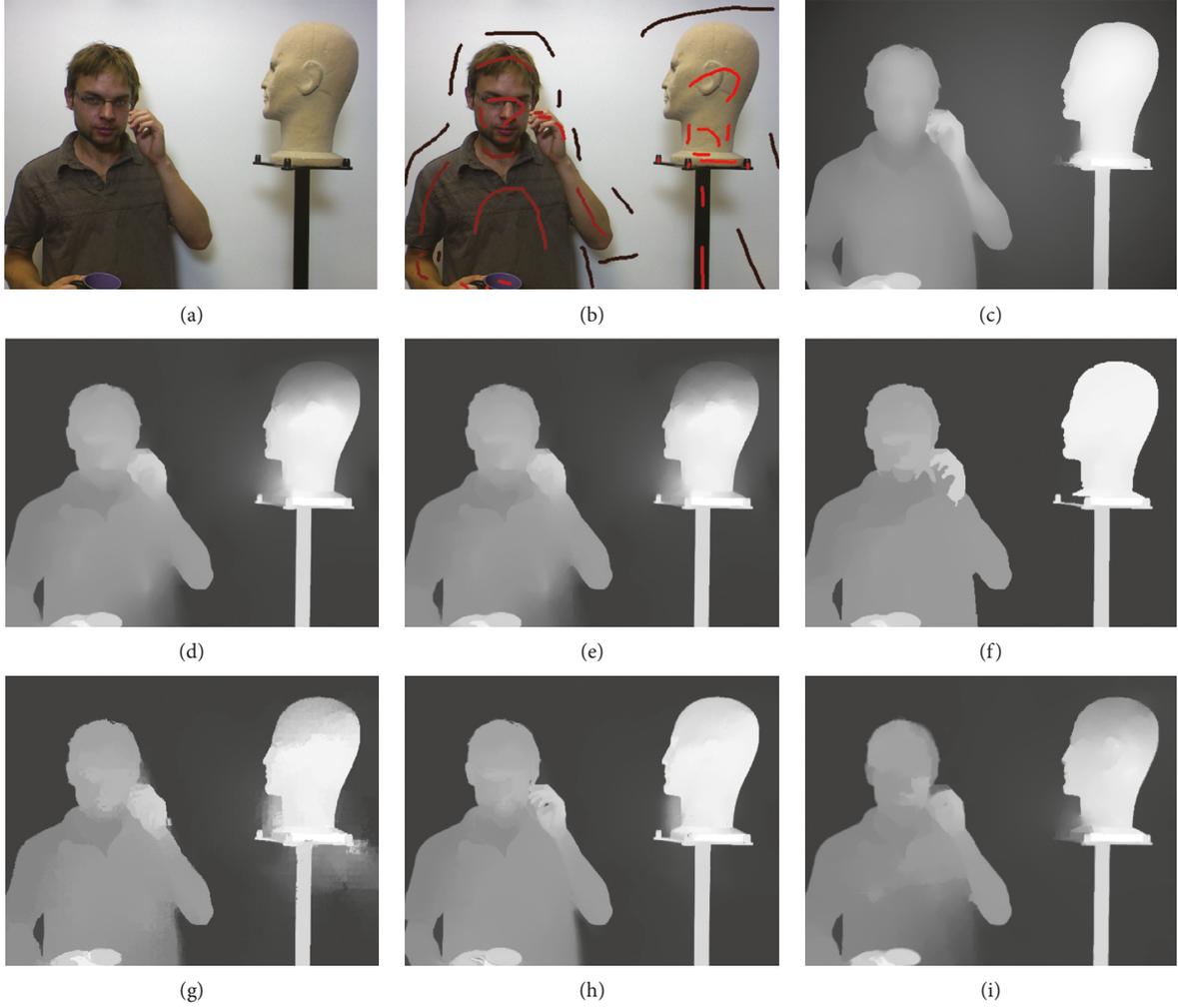


FIGURE 6: Results for RGBZ_05 in the absence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

$n = w \times h$. For each pixel i , x_i , b_i , and $[r_i, c_i]^T$ denote, respectively, the estimated depth value, initial depth value, and spatial coordinate. Here, $b_i = 0$ if $i \in \Omega$; otherwise, b_i is user-assigned depth value. The $n \times 1$ lexicographically ordered column vectors \mathbf{x} and \mathbf{b} are the vector representation of the estimated dense depth-map and initial sparse depth-map, respectively. The objective of 2D-to-3D conversion is to recover \mathbf{x} from \mathbf{b} . This is an ill-posed inverse problem. Local and k -nearest smoothness are introduced to regularize the problem and propose the following energy function:

$$E(\mathbf{x}) = \sum_{i \in \Omega} s_i (x_i - b_i)^2 + \gamma \sum_{i=1}^n \sum_{j \in \mathbf{N}(i)} w_{ij} (x_i - x_j)^2 + \lambda \sum_{i=1}^n \sum_{j \in \mathbf{N}^k(i)} w_{ij} (x_i - x_j)^2, \quad (2)$$

where the first term is data cost, the second term is local smoothness, and the last term is k -nearest smoothness. s_i is the scribble confidence at pixel i obtained from formula (1). $\mathbf{N}^k(i)$ is the set of k -nearest neighbors of pixel i . Feature

vector $\mathbf{f}_i = [\mathbf{I}_i^T, \alpha r_i / \sqrt{h^2 + w^2}, \alpha c_i / \sqrt{h^2 + w^2}]^T$ is used to find k -nearest neighbors. Here, α is a parameter and set as 30 in all experiments. γ and λ in formula (2) are parameters to weigh the local smoothness term and k -nearest smoothness term, respectively. w_{ij} is the Gaussian weight to measure color similarity between pixels i and j and is defined as follows:

$$w_{ij} = \exp\left(-\frac{\|\mathbf{I}_i - \mathbf{I}_j\|_2^2}{2\sigma^2}\right), \quad (3)$$

where σ is the bandwidth parameter and fixed as 0.03 in all experiments.

In formula (2), the data cost is used to measure the consistency between the estimation and user-assigned depth values. Since scribble confidence is incorporated, the proposed method is robust to inaccurate use input. The local smoothness term makes the neighboring pixels with similar colors have similar depth values. To reduce the impact on correct scribbles at texture regions, the k -nearest smoothness term is introduced to make distant pixels with similar features also have similar depth values.

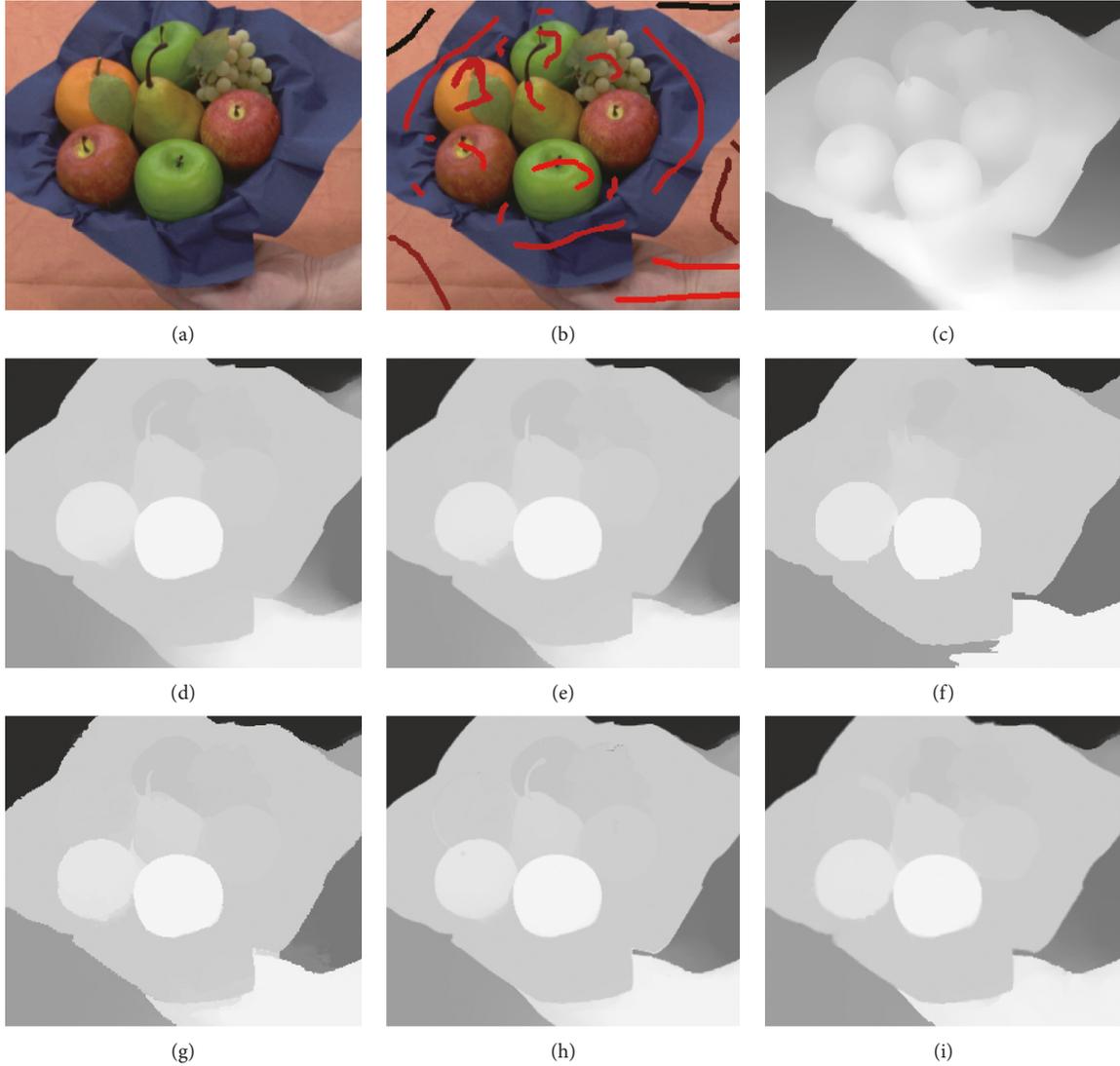


FIGURE 7: Results for RGBZ_07 in the absence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

3.3. Solver. The energy function in formula (2) is minimized to obtain the dense depth-map \mathbf{x} from the sparse depth-map \mathbf{b} . To facilitate computer implementation, formula (2) is rewritten in matrix form as follows:

$$E(\mathbf{x}) = (\mathbf{x} - \mathbf{b})^T \mathbf{S} (\mathbf{x} - \mathbf{b}) + \gamma \mathbf{x}^T \mathbf{L} \mathbf{x} + \lambda \mathbf{x}^T \mathbf{L}^k \mathbf{x}, \quad (4)$$

where \mathbf{S} is an $n \times n$ diagonal matrix whose i -th entry on the diagonal is s_i . \mathbf{L} is the $n \times n$ Laplacian matrix for local neighbors and defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{W} = [w_{ij}]_{n \times n}$ ($j \in \mathbf{N}(i)$) is the $n \times n$ affinity matrix for local neighbors and \mathbf{D} is an $n \times n$ diagonal matrix whose i -th entry on the diagonal is $d_i = \sum_{j \in \mathbf{N}(i)} w_{ij}$. \mathbf{L}^k is the $n \times n$ Laplacian matrix for k -nearest neighbors and defined as $\mathbf{L}^k = \mathbf{D}^k - \mathbf{W}^k$, where $\mathbf{W}^k = [w_{ij}^k]_{n \times n}$ ($j \in \mathbf{N}^k(i)$) is the $n \times n$ affinity matrix for k -nearest neighbors and \mathbf{D}^k is an $n \times n$ diagonal matrix whose i -th entry on the diagonal is $d_i^k = \sum_{j \in \mathbf{N}^k(i)} w_{ij}^k$.

The energy function in formula (4) to be minimized is convex and thus takes its derivative with respect to \mathbf{x} and sets it equal to zero leading to the following system of linear equations:

$$(\mathbf{S} + \gamma \mathbf{L} + \lambda \mathbf{L}^k) \mathbf{x} = \mathbf{S} \mathbf{b}. \quad (5)$$

The equation in formula (5) is sparse and positive definite which means the solution \mathbf{x} can be obtained using the conjugate gradient method.

4. Experimental Results

4.1. Experimental Setup. Four representative test images, RGBZ_01, RGBZ_03, RGBZ_05, and RGBZ_07, from the RGBZ dataset [23] are used to evaluate the performance. The proposed method is compared with several state-of-the-art

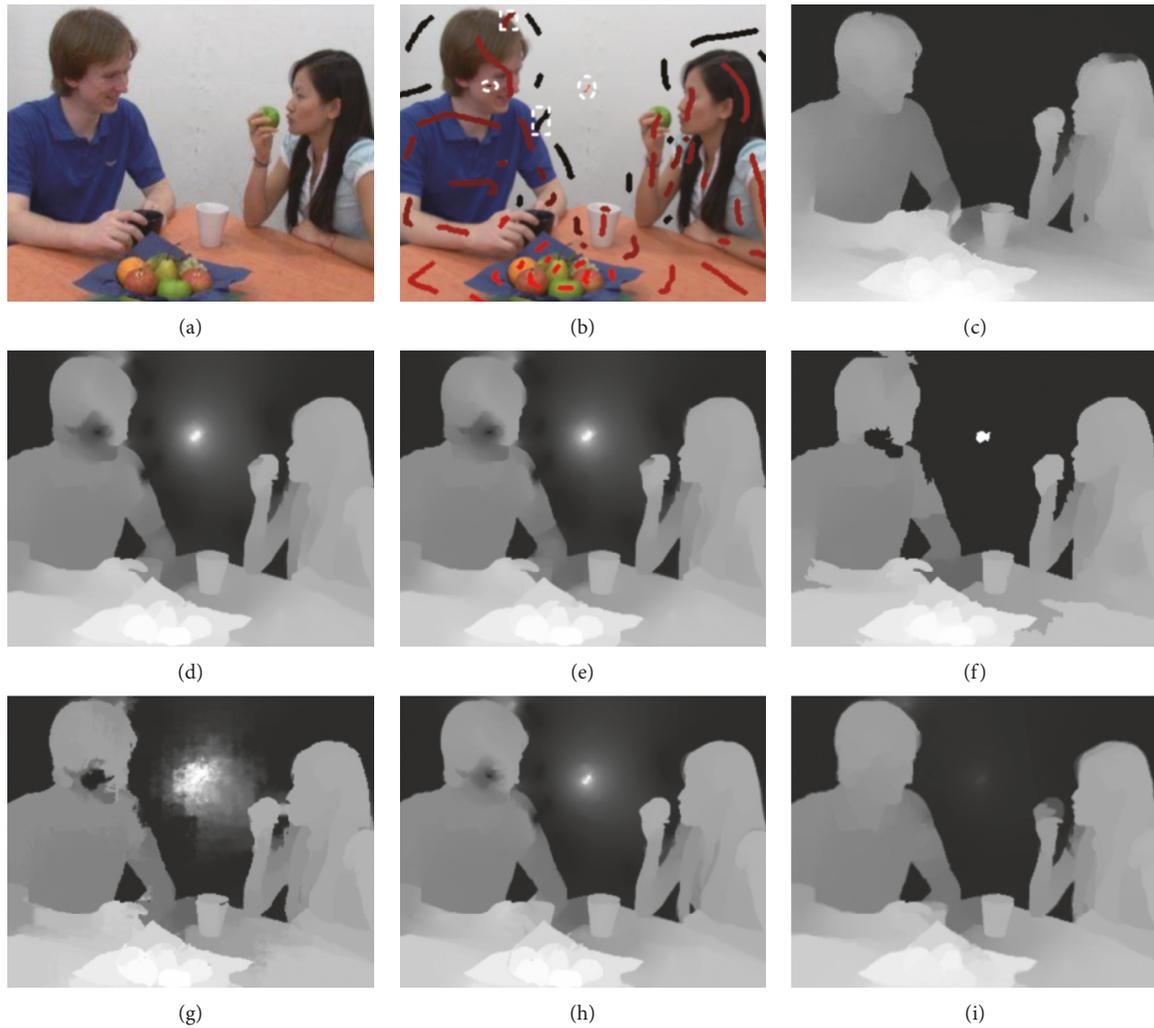


FIGURE 8: Results for RGBZ_01 in the presence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

methods, including RW [4], optimization (OPT) [6], hybrid GC and RW (HGR) [10], superpixel-based optimization (SOPT) [13], and nonlocal RW (NRW) [14]. In the proposed method, the regularization weight parameters γ and λ are fixed to 1 and 10^{-5} , respectively. The local neighbors of formula (1) and (2) are empirically set to 3×3 size square windows centered at each pixel. The parameter k of k -nearest neighbors in (2) is set to 9. Structural similarity (SSIM) and PSNR are used as the quantitative measure for comparison, in which parameters of SSIM are set to default values as suggested by Wang et al. [24].

4.2. Experiments in the Absence of Inaccurate Scribbles. In this section, user input is assumed to be perfectly accurate and show the performance of the proposed method in this case. The SSIM comparison is listed in Table 1. The PSNR comparison is shown in Table 2. As shown in Tables 1 and 2, the proposed method is comparable to current optimization method [6] when user input is accurate. Figures 4–7 show qualitative comparisons for different methods on the four test

images. It can be seen that the proposed method is superior in reducing depth bleeding artifacts compared with the previous optimization method [6]. The reason is that the k -nearest smoothness term in formula (2) is effective in preserving sharp depth boundaries [14]. In summary, the proposed method can be safely used in the case of accurate input.

4.3. Experiments in the Presence of Inaccurate Scribbles.

In this section, some inaccurate scribbles are added on the abovementioned experiments by roughly drawing labels across some randomly selected object boundaries (see regions within white squares of Figures 8(b)–11(b)) or wrongly drawing inside randomly selected regions (see regions within white circles of Figures 8(b)–11(b)).

The SSIM comparison in this case is listed in Table 3, and PSNR comparison is shown in Table 4. It can be seen that the proposed method is superior to other approaches when inaccurate input is present, and obtains the highest SSIM and PSNR values in average. This shows that scribble confidence can help resist inaccurate scribbles. The

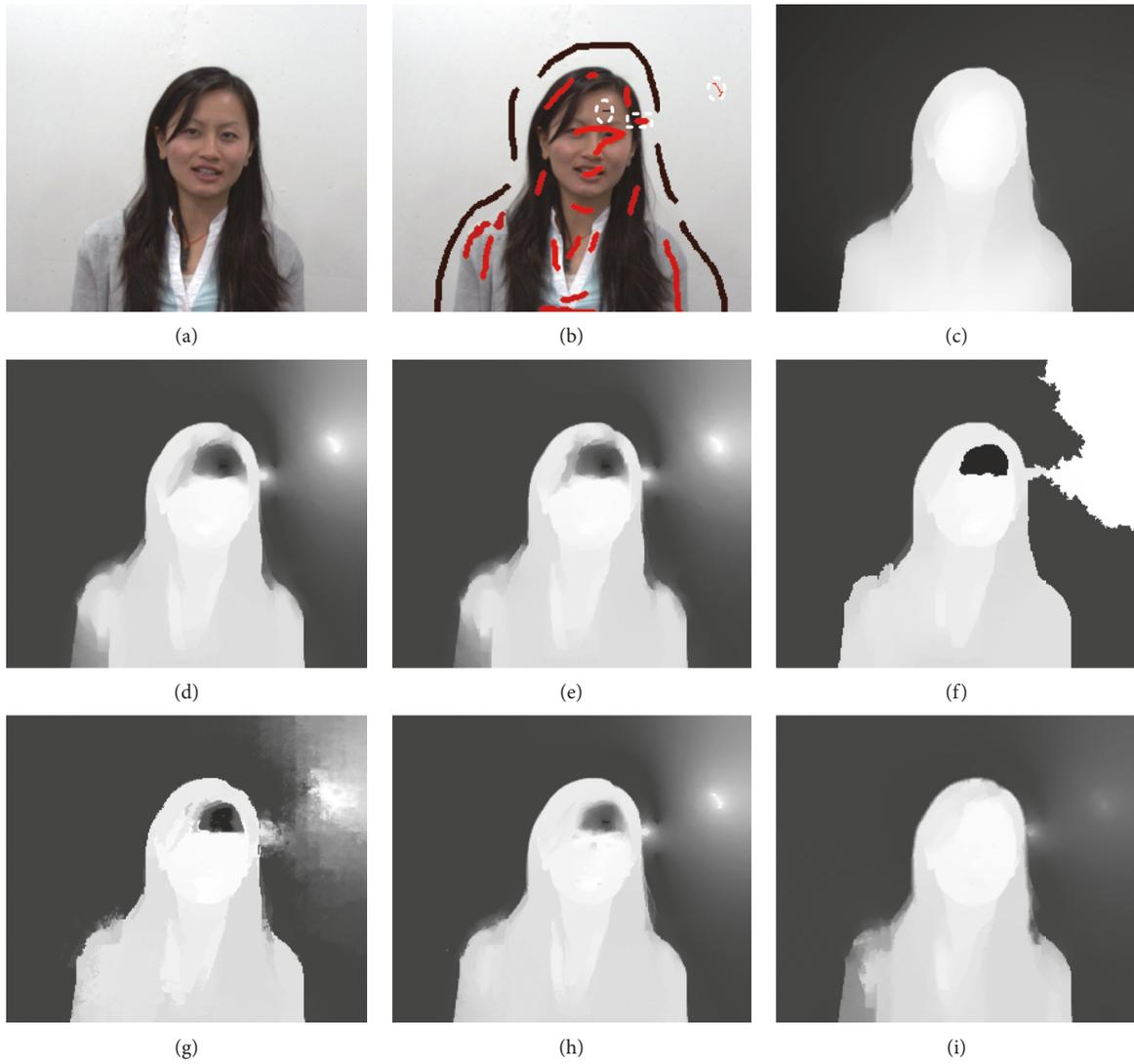


FIGURE 9: Results for RGBZ_03 in the presence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

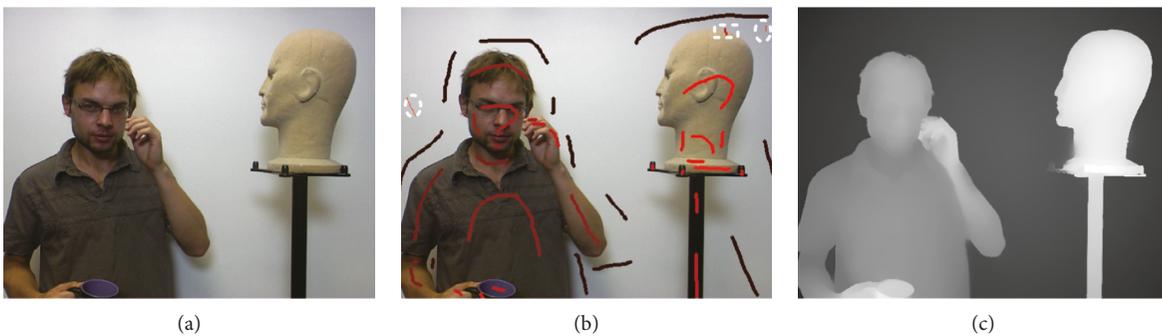


FIGURE 10: Continued.

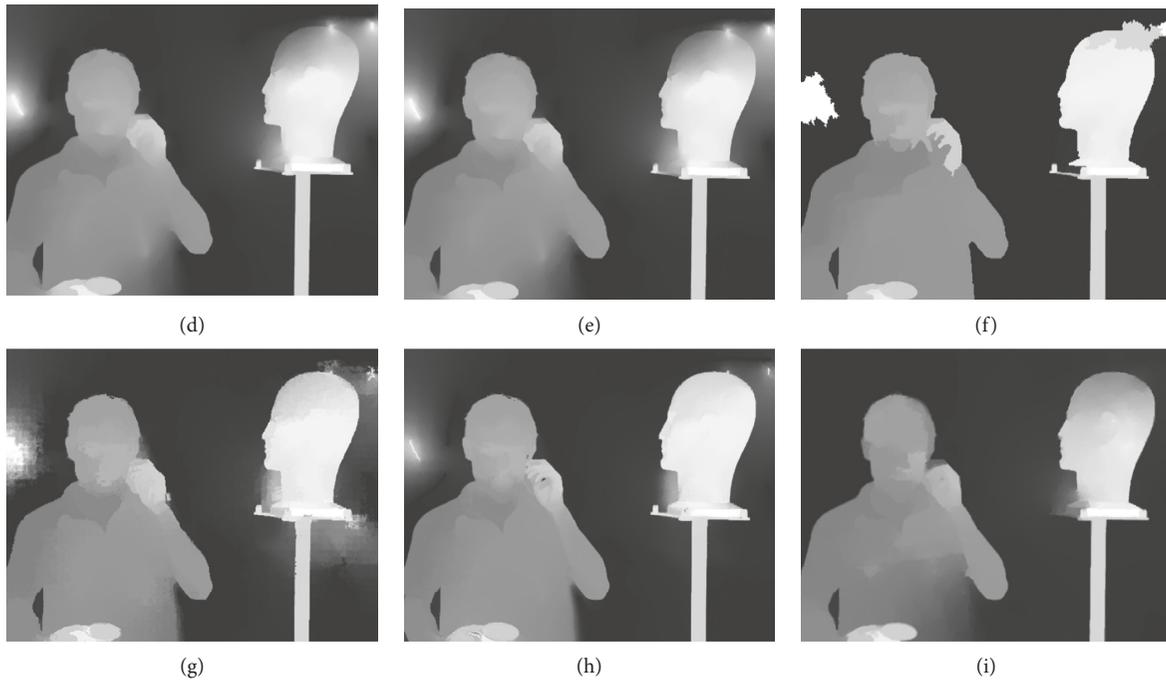


FIGURE 10: Results for RGBZ_05 in the presence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

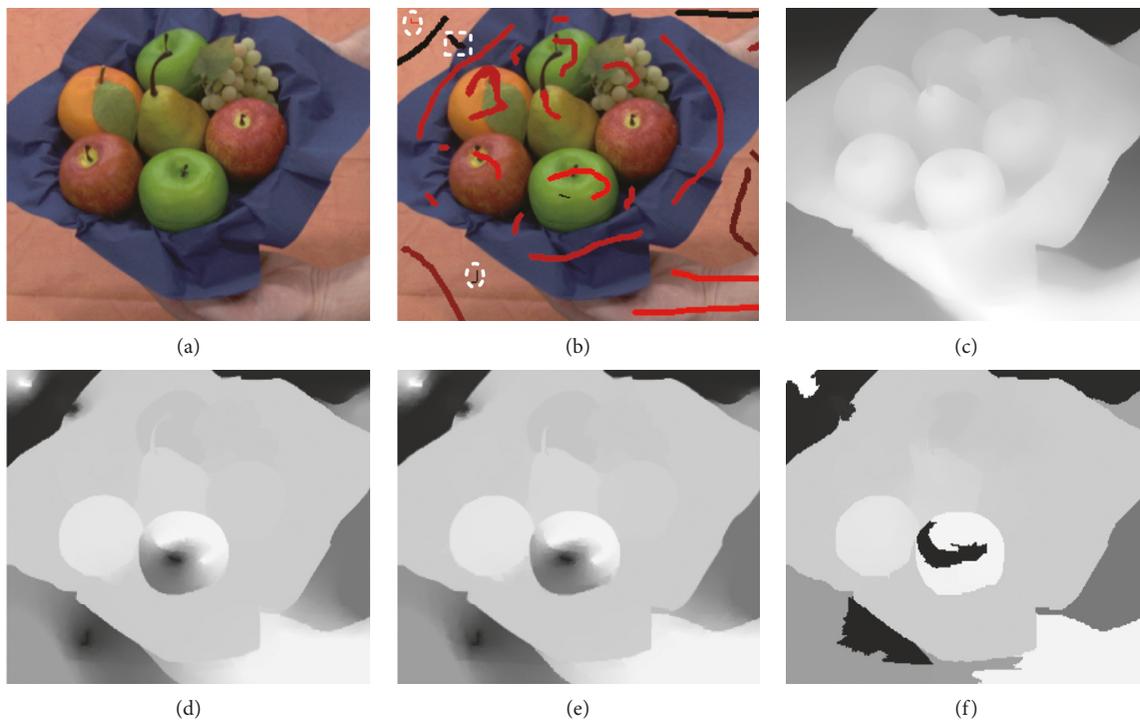


FIGURE 11: Continued.

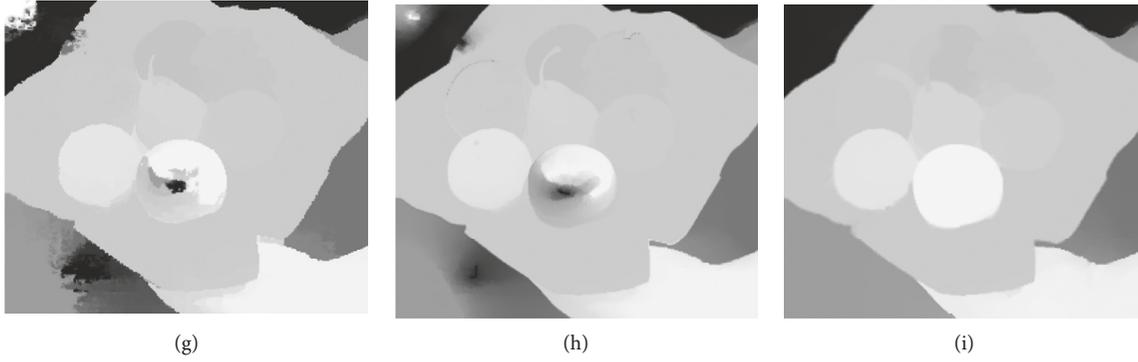


FIGURE 11: Results for RGBZ_07 in the presence of inaccurate scribbles. (a) Input image. (b) User scribbles. (c) Groundtruth. (d) Depth of RW [4]. (e) Depth of OPT [6]. (f) Depth of HGR [10]. (g) Depth of SOPT [13]. (h) Depth of NRW [14]. (i) Depth of proposed.

TABLE 3: SSIM comparison in the presence of inaccurate scribbles.

Image	Method					
	RW [4]	OPT [6]	HGR [10]	SOPT [13]	NRW [14]	Proposed
RGBZ_01	0.77	0.77	0.88	0.76	0.78	0.85
RGBZ_03	0.76	0.76	0.77	0.74	0.79	0.83
RGBZ_05	0.90	0.89	0.89	0.86	0.90	0.92
RGBZ_07	0.86	0.86	0.83	0.84	0.86	0.91
Average	0.82	0.82	0.84	0.80	0.83	0.88

The best SSIM at each row is shown in bold.

TABLE 4: PSNR comparison in the presence of inaccurate scribbles.

Image	Method					
	RW [4]	OPT [6]	HGR [10]	SOPT [13]	NRW [14]	Proposed
RGBZ_01	19.5	19.5	21.0	17.2	20.6	24.1
RGBZ_03	14.4	14.3	10.3	13.9	15.5	19.0
RGBZ_05	18.0	17.7	17.2	18.0	20.3	21.3
RGBZ_07	17.2	17.2	15.0	16.8	18.5	22.3
Average	17.3	17.2	15.9	16.5	18.7	21.7

The best SSIM at each row is shown in bold.

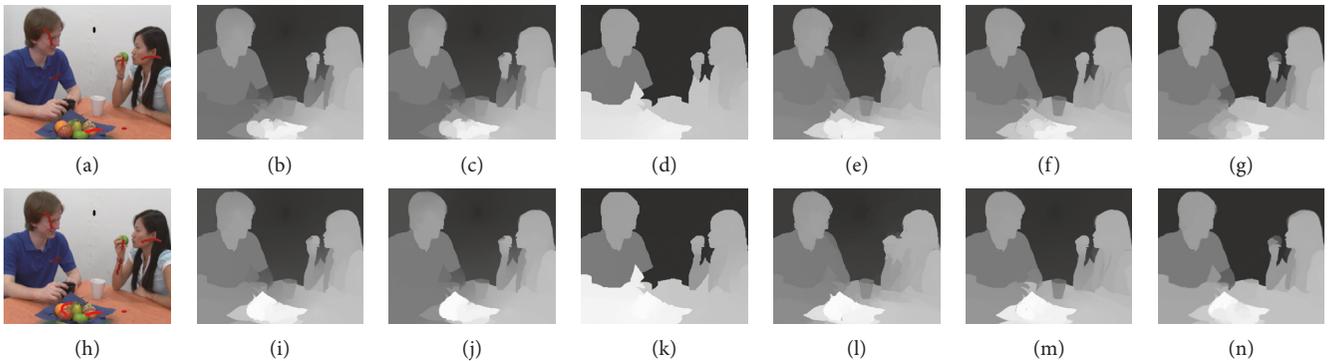


FIGURE 12: Results of depth for RGBZ_01 with sparse input. (a) User scribbles 1. (b) Depth of RW for scribbles 1. (c) Depth of OPT for scribbles 1. (d) Depth of HGR for scribbles 1. (e) Depth of SOPT for scribbles 1. (f) Depth of NRW for scribbles 1. (g) Depth of proposed for scribbles 1. (h) User scribbles 2. (i) Depth of RW for scribbles 2. (j) Depth of OPT for scribbles 2. (k) Depth of HGR for scribbles 2. (l) Depth of SOPT for scribbles 2. (m) Depth of NRW for scribbles 2. (n) Depth of proposed for scribbles 2.

performance of current methods degrades significantly in the case of inaccurate input, since they assume all scribbles are perfectly accurate.

The qualitative comparisons for different methods are shown in Figures 8–11. Current methods generate undesirable visual artifacts around inaccurate labeled regions, as shown in Figures 8–11. Thanks to the scribble confidence, the proposed method successfully reduces these artifacts caused by inaccurate input.

4.4. Experiments for Sparse Labeling. In this section, user input is assumed to be very sparse. In the first row of Figure 12, results are shown for a very sparse input which only contains seven strokes. In the second row, two strokes are added to the input image. It can be seen that the performance of all methods is improved as increase in the number of accurate scribbles. NRW and the proposed method are superior to others in preserving depth discontinuities since they both use nonlocal regularization. As analyzed in Section 3.1, although the proposed method may mistake correct labels for incorrect ones when they are located in texture regions, the proposed method can obtain acceptable results even with very sparse scribbles thanks to the k -nearest smoothness in formula (2), as shown in Figures 12(g) and 12(n).

5. Conclusion

Semiautomatic 2D-to-3D conversion has proven to be an effective solution for alleviating 3D content shortage. The key is sparse-to-dense depth conversion from user scribbles. Existing methods assume user input is entirely accurate, and even small errors may degrade the depth quality dramatically. To alleviate this problem, color difference between labeled pixels and neighbors is used to compute scribble confidence, and a confident optimization method is proposed for sparse-to-dense depth conversion. Furthermore, k -nearest smoothness is introduced to make the proposed method perform well even with very sparse input. The experiments demonstrate that the proposed method is superior to existing methods when inaccurate input is present, while at the same time competitive results are obtained when all scribbles are accurate.

Currently, the proposed method mainly focuses on 2D-to-3D conversion for images. In future, the proposed method will extend to videos.

Conflicts of Interest

The author declares that there are no conflicts of interest in the publication of this article.

Acknowledgments

This research was supported by the Zhejiang Provincial Natural Science Foundation of China (LY16F010014 and LY18F020025), the Ningbo City Natural Science Foundation of China (2017A610109), and the Educational Commission of Zhejiang Province of China (Y201533511).

References

- [1] J. Chen and M. Huang, "2D-to-3D conversion system using depth map enhancement," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 3, pp. 1159–1181, 2016.
- [2] W. Huang, X. Cao, K. Lu, Q. Dai, and A. C. Bovik, "Toward naturalistic 2D-to-3D conversion," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 724–733, 2015.
- [3] A. Kanchan and T. Mathur, "Recent trends in 2D to 3D image conversion: algorithm at a glance," *International Research Journal of Engineering and Technology*, vol. 4, no. 4, pp. 3480–3484, 2017.
- [4] R. Rzeszutek, R. Phan, and D. Androustos, "Semi-automatic synthetic depth map generation for video using random walks," in *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1–6, Barcelona, Spain, July 2011.
- [5] R. Rzeszutek and D. Androustos, "Label propagation through edge-preserving filters," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, pp. 599–603, Florence, Italy, 2014.
- [6] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M. Gross, "Stereobrush interactive 2D to 3D conversion using discontinuous warps," in *Proceedings of Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pp. 47–54, Vancouver, Canada, 2011.
- [7] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, pp. 2024–2039, 2016.
- [8] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proceedings of IEEE International Conference on Computer Vision*, pp. 2650–2658, Santiago, Chile, December 2015.
- [9] J. Xie, R. Girshick, and A. Farhadi, "Deep3d: fully automatic 2D-to-3D video conversion with deep convolutional neural networks," in *Proceedings of European Conference on Computer Vision*, pp. 842–857, Amsterdam, Netherlands, October 2016.
- [10] R. Phan and D. Androustos, "Robust semi-automatic depth map generation in unconstrained images and video sequences for 2D to stereoscopic 3D conversion," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 122–136, 2014.
- [11] R. Rzeszutek and D. Androustos, "Propagating sparse labels through edge-aware filters," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 17–24, 2015.
- [12] S. Iizuka, Y. Endo, Y. Kanamori, J. Mitani, and Y. Fukui, "Efficient depth propagation for constructing a layered depth image from a single image," *Computer Graphics Forum*, vol. 33, no. 7, pp. 279–288, 2014.
- [13] S. Wu, H. Yuan, P. An, and P. Cheng, "Semi-automatic 2D-to-3D conversion using soft segmentation constrained edge-aware interpolation," *ACTA Electronica SINICA*, vol. 43, no. 11, pp. 2218–2224, 2015.
- [14] H. Yuan, S. Wu, P. Cheng, P. An, and S. Bao, "Nonlocal random walks algorithm for semi-automatic 2D-to-3D image conversion," *IEEE Signal Processing Letters*, vol. 22, no. 3, pp. 371–374, 2015.
- [15] Z. Liang and J. Shen, "Consistent 2D-to-3D video conversion using spatial-temporal nonlocal random walks," in *Proceedings of IEEE International Conference on Digital Signal Processing*, pp. 672–675, Beijing, China, October 2016.
- [16] A. Lopez, E. Garces, and D. Gutierrez, "Depth from a single image through user interaction," in *Proceedings of Spanish*

- Computer Graphics Conference*, pp. 1–10, Zaragoza, Spain, 2014.
- [17] L. Vosters and G. D. Haan, “Efficient and stable sparse-to-dense conversion for automatic 2-D to 3-D conversion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 3, pp. 373–386, 2013.
- [18] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, “Epicflow: edge-preserving interpolation of correspondences for optical flow,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1164–1172, Boston, MA, USA, 2015.
- [19] E. Zemene and M. Pelillo, “Interactive image segmentation using constrained dominant sets,” in *Proceedings of European Conference on Computer Vision*, pp. 278–294, Amsterdam, Netherlands, October 2016.
- [20] C. Oh, B. Ham, and K. Sohn, “Robust interactive image segmentation using structure-aware labeling,” *Expert Systems with Applications*, vol. 79, pp. 90–100, 2017.
- [21] K. Subr, S. Paris, C. Soler, and J. Kautz, “Accurate binary image selection from inaccurate user input,” *Computer Graphics Forum*, vol. 32, no. 2, pp. 41–50, 2013.
- [22] S. Hillaire, A. Lecuyer, T. Regiacorte, R. Cozot, J. Royan, and G. Breton, “Design and application of real-time visual attention model for the exploration of 3D virtual environments,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 3, pp. 356–368, 2012.
- [23] S. Lu, X. Ren, and F. Liu, “Depth enhancement via low-rank matrix completion,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3390–3397, Columbus, OH, USA, 2014.
- [24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.



Hindawi

Submit your manuscripts at
www.hindawi.com

